Macro-Operations for Hypertext Construction[1]

Rainer Hammwöhner
Universität Konstanz
Fachgruppe Informationswissenschaft
Projekt TWRM-TOPOGRAPHIC
Postfach 5560 D-7750 Konstanz

Abstract

This paper deals with special aspects of the automatic construction of hypertexts from journal articles as a means to improve the access methods to full text information. In the first section of the paper a text linguistic approach to hypertext construction is introduced and related to the state of the art in information retrieval technology. The next section gives a notion of global structuring and the appropriate granularity of hypertexts based on text linguistic evidence. A formalization of the resulting hypertext model is outlined in the final section.

I Introduction

One of the major shortcomings of current full text information systems[2] is that almost the whole effort is spent into the support of query formulation. The user of information retrieval systems — being in an "anomalous state of knowledge" (Belkin et al. 82) — has problems to express his informational needs, therefore query formulation is supported by expert systems, natural language interfaces[3], (eg. Biswas et al. 87a, 87b) and user models (Brajnik et al. 87). Having retrieved a set of relevant texts from a database the user has to face the following problems:

- Based on the best match paradigm (Robertson 80), most information retrieval systems measure the relevance of a document according to the degree of similarity between document representations (eg indexes) and query (search terms). A user being interested in pieces of information scattered over several texts may therefore have to read couples of redundant texts before he reaches a text which contains a new informational item.
- Additionally the presentation techniques employed in conventional retrieval systems are rather poor. In the worst case the user is confronted with a list of references.

The difficulties to obtain information from a set of retrieved texts lead to draw-backs in full text retrieval performance (Blair 80, Blair/Maron 85, Tenopir 85).

[1] This is an enhanced version of a paper published in: Jonassen, D. (ed): Proceedings NATO Advanced Research Workshop "Designing Hyper-text/Hypermedia for Learning". Springer, 1990, 71-95.

[2] An overview of the state of the art in information retrieval gives Belkin/Croft 87.

[3] The role of AI methods in information retrieval is outlined by Smith 87.

One way to overcome these shortcomings — the best one for users with fact oriented questions — is to complement full-text retrieval by text-based question answering systems[4] (Rau 87a,b, Simmons 87). Relevant chunks of information are chosen according to an inference process, which additionally comprises the relevant information to a concise answer. In this paper we will deal with an alternative approach to information retrieval and presentation.

A common feature of various recently developed information systems is the decomposition of linear document structure which is enforced by conventional print media. Instead, an organization (networks or hierarchies) of information units of different forms (textual, graphical and pictorial presentation modes may be combined) is provided. Additionally the presentation of textual information is enhanced by alternative presentation styles like tables etc. (Stibic 85). Documents organized this way are called *hypertexts* (an overview gives Conklin 87). Hypertext systems are devoted to the exploratory paradigm (Bates 85) — reading a hypertext means traversing a network of text-units using a browsing-facility. Additionally hypertext systems offer string oriented retrieval functions[5], hierarchies of organizational text units like the *tocs* of Trigg's textnet (Trigg/Weiser 86) or predefined paths (textnet) to support hypertext navigation. The conversion of texts into hypertext with its variety in presentation and navigation techniques is an alternative approach to improve online retrieval performance.

The different approaches to the conversion of texts into hypertext can be distinguished with respect to the answers given to the following two questions:

- What are the text units constituting a hypertext?
- What sort of links between the units will be provided?

The I^{3}R-system (Croft/Thompson 87) for instance is based on statistical clustering. Hypertext-units are references to documents and links are based on a similarity measure or on citation. Another system based on clustering is the lOTA-system (Defude/ Chiamarella 87) which allows for indexing and therefore interrelation of parts of documents (chapters etc.). An adaptation of this statistical approach to hypertext is given by Larson (Larson 88). Frisse (Frisse 88) on the other hand proposes a semi-automatic text decomposition method. Text units of the resulting hypertext are identical with passages of the linear text. The hierarchical organization of the hypertext is based on the structure of the original text (e.g. the text divisions like chapters or sections)[6]. Non-hierarchical semantic or rhetoric links between text units may be provided by the user.

The approach we propose[7] is based neither on statistical evidence nor on the surface structure of texts, but on text-linguistic regularities. Many semantic theories of text, like text grammars based on macrostructures (eg van Dijk 80b) or the definition of semantic coherence through binary relations (Hobbs 83, Mann/Thompson 88) are based on the two-dimensional structure of text (Gülich/Raible 77 pp. 51-55). Based on these text-linguistic models texts may be converted to hypertext as follows:

[4] Question answering systems and retrieval systems employ different notions of relevance, which in one case is based on informational needs and in the other on thematic overlap (Swanson 77).

[5] Context free retrieval functions don't seem to be sufficient to the heavily content dependent navigation model of hypertext systems, therefore Frisse (Frisse 87) proposes context sensitive (but still string oriented) retrieval functions.

[6] Similar to hypertext systems, which originally emerged from text formatters – eg. Superbook (Remde et al. 87).

[7] A partial implementation of the hypertext model as proposed in this paper is the TWRM-TOPOGRAPHIC system, which was developed at the University of Constance from 1982-1988 (Thiel/Hammwöhner 86).

The text is fragmented to coherent text units which are mapped to semantic representations (Hahn/Reimer 86).

- Content oriented relations similar to the rhetoric relations mentioned above can be computed on the ground of these representations (Hammwöhner/Thiel 87). Thus, text passages taken from linear text can be rearranged as networks, such that every possible path in this network is semantically coherent.
- Macrostructures which resemble the hierarchical structures as found in conventional hypertexts (content nodes) reflect the topical structure of documents.

Macrostructures can be computed either

- bottom up for the purpose of text analysis (the text — the leaves of the macrostructure — is given): so called macrorules are applied to text units creating a more general topical description (van Dijk 80b) — or
- top down for the purpose of text generation (the topic — the root of the macrostructure — is given): the inverse rules are used to create more specific propositions (Garcia-Berrio/Mayordomo 88).

Hypertext systems comprise both of these aspects. A hypertext may be regarded as a static network with a given hierarchical deep structure (comparable to the text-graphs as introduced by Reimer/Hahn 88 for linear texts). Navigating a hypertext implies the choice of sub-graphs from the network which fit the user's (topical) interest. The application of macro-rules with respect to this interest[8] may be viewed as a user oriented reinterpretation of the hypertext (bottom up) as well as a navigation driven text construction process (top down). Macro-rules implement a context oriented notion of relevance (Tiamiyu/ Ajiferuke 88) fulfilling the maxims of relation and quantity (Grice 75) which deal with the choice of relevant and the elimination of redundant information. This paper will give a semi-formal description of macrorules for the construction of hypertext deep structures (based on a frame like formalism) and outline their role in hypertext navigation.

The application of macro-rules is controlled by prototypical text plans, comparable to superstructures (van Dijk 80a) or story grammars (Rumelhart 75). The possibility to compare the actual state of navigation with a text plan and as a result provide the user with discourse cues as demanded by Charney 87 will help to avoid confusion in hypertext navigation as observed by Jones 87[9].

2 Macrostructures

The oldest science which deals with structuring and formulation of text is rhetoric which has a tradition reaching back to ancient Greece. The special fields of rhetoric traditionally are:

- inventio: the discovering of the very ideas which shall be expressed in a text,
- dispositio: the ordering of these ideas and
- elucutio: the finding of adequate formulations.

[8] One of the macro-rules for instance describes the deletion of accidental information. The decision, which information shall be regarded as accidental, should consider whether it is in the scope of the user’s interest or not.

[9] Kieras 82 shows and explains the effect of discourse cues – especially cues on correct generalizations – on the comprehension of simple technical prose according to a text model based on macro-structures.

Writing a text may be thought of as a interlocking process of inventio, dispositio and elucutio which leads to a stepwise refinement. A (still to be developed) rhetoric of hypertext has to reconsider especially the role of text disposition, which takes place in two phases. The hypertext author provides a network structure (dispositio) which contains chunked pieces of information (inventio) verbally expressed in text units (elucutio). The final ordering of these text units to more or less linear hypertext paths is done by the reader of the hypertext. In the search for hypertext disposition rules it should be helpful to consider theories of rhetorics and text linguistics about the structuring of linear text.

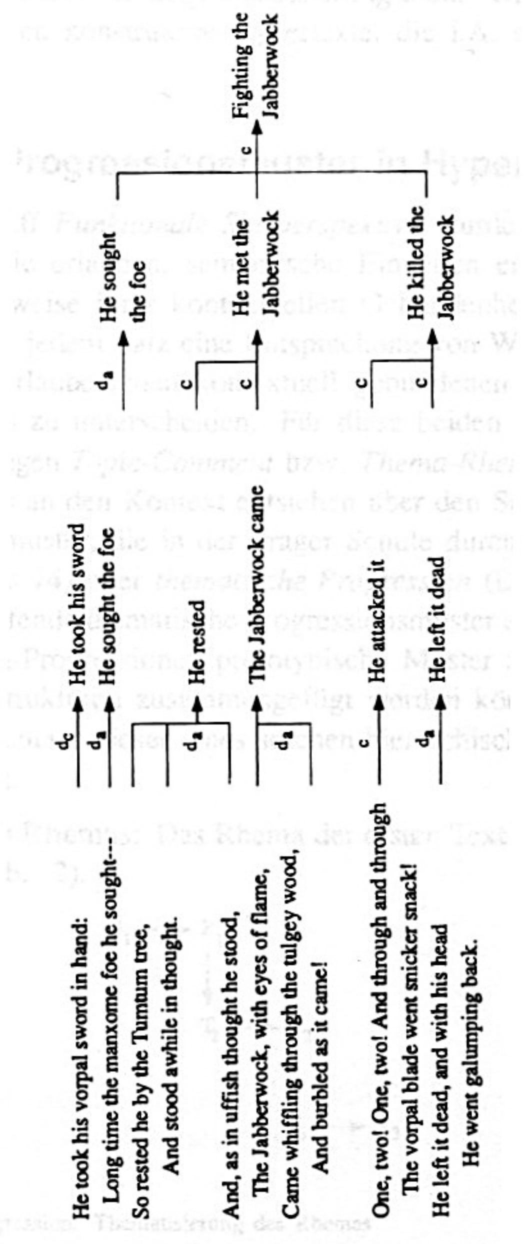
He took his vorpal sword in hand:
Long time the manxome foe he sought—
So rested he by the Tumtum tree,
And stood awhile in thought.
And, as in uffish thought he stood,
The Jabberwock, with eyes of flame,
Came whiffling through the tulgey wood,
And burbled as it came!
One, two! One, two! And through and through
The vorpal blade went snicker snack!
He left it dead, and with his head
He went galumping back.
dc
He took his sword
da
He sought the foe
da
He rested
da
The Jabberwock came
c
He attacked it
da
He left it dead
da
He sought the foe
c
c
He met the Jabberwock
c
c
He killed the Jabberwock
c
Fighting the Jabberwock

Figure 1 The (slightly simplified) macrostructure of three verses taken from the poem „Jabberwocky" (Carroll 39, pp 140-142).

2.1 Macrostructures of linear text

Garcia-Berrio/ Mayordomo 88 point out the strong connection between the rhetoric concept of dispositio and the text-linguistic[10] notion of macrostructure as developed by van Dijk (van Dijk 80a, 80b, Ballmer 76). Macrostructures reaching beyond the domain of single sentences are elements of the semantic deep structure (a hierarchy) of texts. Macrostructures are sequences of propositions, which can be expressed in first order calculus. They can be derived from the microstructure of a text (surface structure) by the application of so called macro-rules. These macro-rules are semantic transformation rules mapping tuples of propositions to more general (macro-)propositions. Applied recursively, these transformations produce more and more general descriptions of the text. Thus, a hierarchy can be built which reaches from sentence topics over paragraph and chapter topics to the topic of the whole text[11]. The four macro-rules which are employed in this process are:

I. The first macro-rule deals with the deletion of accidental information. Information which is not needed to understand the subsequent text is deleted:

   Attacking from behind the Tumtum tree the knight killed the Jabberwocky. $\xrightarrow{d_a}$ *The knight killed the Jabberwocky*[12].

II. The application of the second macro-rule results in the deletion of constitutional information. Information which can be inferred by presupposition is deleted:

   He fought the Jubjub bird and didn't shun the frumious Bander-snatch. He is a brave man. $\xrightarrow{d_c}$ *He fought the Jubjub bird and didn't shun the frumious Bandersnatch.*

III. The third macro-rule treats simple generalization: special information is re-placed by more general information:

   Alice was having smalltalk with a tiger-lily, a rose and some daisies. $\xrightarrow{g}$ *Alice was having smalltalk with some garden flowers.*

IV. The fourth macro-rule is about the construction of propositions: A proposition is constructed which comprises a set of propositions from the text or macro-propositions from macrostructures:

   An egg with hands and feet is sitting on the wall. $\xrightarrow{c}$ *Humpty Dumpty is sitting on the wall.*

Figure 1 shows a partial macrostructure of a poem by Lewis Carroll. It can be discerned in spite of new invented nonsense words, which are spread all over the text (according to Burchfield 76 for instance "manxome" or "vorpal"). Although the reader is furnished with a fixed set of macro-rules, the derived macrostructure is not independent from the readers interests (if the reader had any interest in the "tulgey wood", it would possibly appear in a high level macrostructure) and knowledge (eg that a sword is usually taken in hand).

[10] A general introduction to text-linguistic is provided by de Beaugrande/Dressler 81.

[11] Van Dijk’s model represents one of the well elaborated text models originating from transformal-generative grammar. The other theory that must be mentioned in this context is Petöfi’s Text-Structure Word-Structure Theory (TeSWeST, Petöfi 79). Similar to van Dijk’s micro- and macro-structure Petöfi distinguishes between linear text manifestation and text basis. Both text basis and macrostructure describe a global text structure derived by semantic transformation rules. TeSWeST employs a large formal apparatus, which makes it more difficult to adapt to new applications than the notion of macro-structures with its more informal treatment.

[12] The examples are adapted from Carroll 39

2.2 What is in a text unit?

Pivotal point of global coherence in hypertext are the text units. The sub-structures of text units are unchangeable during dialog, whereas the position of a text unit in a hypertext path is fixed only to that extent, that its predecessor and successor must be linked to it. This uncertainty in the final positioning of a text unit demands that text units must be self sufficient with respect to the following aspects:

- thematic unity: the boundaries of a text unit must coincide with the boundaries of a semantic theme.
- anaphora, pronouns: all occurring anaphora and pronouns must be resolvable within the text unit itself or within the preceding hypertext path.

Taken into consideration that we want to construct hypertexts from sets of linear text by a knowledge based text decomposition, the question arises, what text parts fulfill these conditions. A well-founded answer to this question will probably be helpful for the modularization of hypertexts by human authors as well — a discussion which is often dominated by the idiosyncrasies of contemporary hypertext systems (eg. Conklin 87 p. 42).

The role of text units in hypertext is comparable to that of paragraphs in linguistics, both bridge a gap between structural levels: fixed ordering — flexible ordering in the case of text units, sentence level — text level in the case of paragraphs. Both of these contrasting pairs reflect the difference between global and local coherence phenomena. Although the paragraph is not generally believed to be a canonical text segmentation unit — Phillips 85 for instance refers to the frequency of erroneous paragraphing (p. 90) — the importance of the paragraph for text structuring is nevertheless widely accepted in linguistics (Longacre 79) and psychology (Stark 88, Garnes 87, Koen et al. 69).

Paragraph is considered as semantic unity discussing a certain discourse topic, (Garcia-Berrio/Mayordomo 88, Longrace 79, Pike/Pike 77), thus a paragraph oriented decomposition of text can't be based on syntactic evidence (eg. indentation) alone, but must ground on a semantic model of text, which is capable of dealing with incorrect syntactic paragraphing. (Stark 88 shows that human readers correctly distinguish semantic paragraphs in spite of erroneous paragraph markers.) Furthermore there is no anaphoric reference to the paragraph topic from outside the paragraph (Giora 83). Thus, the prerequisites for text units as mentioned above are fulfilled.

An additional property of the paragraph, which makes it a fruitful concept for hypertext, is the typing of paragraphs[13], according to their internal structure (linkage between sentences, local coherence), thematic progression (Daneš 78) and discourse function. Narrative paragraphs with temporal linking and backward referenced can be distinguished from expository paragraphs with causal linking (Longacre 74, 76, Zimmermann 78)[14]. Typed paragraphs can be regarded as the terminal symbols of a hypertext grammar based on macro- and superstructures (van Dijk 80a).

[13] An overview on text typology is provided by Große 74.

[14] A restriction of the paragraph types available in a hypertext is a means to reduce the complexity of text analysis and hypertext planning. In this paper we will further deal with descriptive or expository paragraphs with a-temporal linking. Therefore we will not encounter problems like procedural plausibility etc. which are typical for narrative (temporal linking) paragraphs.

Additionally to the types of discourse two types of thematic bordering can be discerned (Giora 83):

I. The paragraph is cut off *before* a new discourse topic is introduced (eg the second verse of the poem shown in figure 1).

II. The paragraph is cut off *immediately after* the new discourse topic is presented, providing a stronger link to the following paragraph (eg the first verse of the poem shown in figure 1).

2.3 Macrostructures in hypertext ?

Macrostructures in hypertexts differ in several aspects from macrostructures of linear texts. In linear texts macrostructures give — as mentioned above — a hierarchical representation of the text's topical structure on a meta level. They reflect the process of generalization in text analysis or stepwise refinement in text generation. Both of these processes are ruled by a special notion of relevance (eg. before the application of the first macro-rule it must be determined which information can be regarded as accidental and thus be deleted). Where text generation is driven by contextual relevance — what ideas does the author want to express —, text analysis is guided primarily by textual relevance — is a particular information important to understand the following text (but also does it fit the reader's information needs) (van Dijk 79).

Macrostructures *of* hypertexts as a meta level description of the deep structure of the complete hypertext network would be the exact correspondence to macrostructures of linear text — a notion of hypertext macrostructure we don't want to consider any further. Hypertext navigation may be understood as the construction of texts — so called hypertext paths — which are built up from a set of given pieces — i.e. the text units — but nevertheless fulfill the criteria of textual wellformedness and therefore have macrostructures themselves. Elements of these path-macrostructure are macro-text-units which are derived from the original text units by the application of macro-operations[15]. These macrostructures *in* hypertext comprise the (contextual and textual) relevant information of a set of text units in a condensed form[16]. These text units may stem from different linear texts, therefore macrostructures in hypertext reflect special aspect of intertextuality of document fragments (Begthol 86). Regarded as part of the hypertext derived text units can be used as navigational aids in hypertext browsing (the presentation[17] of macro-text-units for instance can help to choose the appropriate hypertext path to follow).

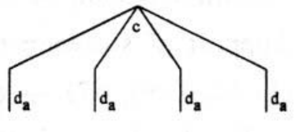

Figure 2 Several text units describe aspects of a single concept

[15] In the following we distinguish macro-operations from macro-rules which were defined on linear texts.

[16] Obviously there is a relation to automatic text summarization (for an overview see Hutchins 87) and especially to text condensation as developed in the TOPIC-Project (Reimer/Hahn 87). The notion of hypertext macro-structure may be viewed as a generalization of the TOPIC text-graphs. The use of interest profiles is similar to the Susy-System (Fum et al. 82).

[17] The content of a macro-text-unit may be graphically presented as a conceptual network (Thiel/Hammwöhner 86) or as a textual abstract (Sonnenberger 88).

The definition of prototypical macrostructures (textplans) helps to adapt the dialog to special informational needs and user purposes. A hypertext containing text fragments taken from computer magazines (a domain all further examples are taken from) can be traversed for instance following a path:

- which gives a description of one special device (fig. 2)
- which compares two (or several) devices with respect to their properties (fig. 3).

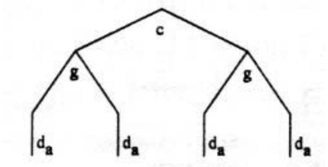

Figure 3 Instances of a generic concept are compared with respect to their properties

3 A frame-oriented hypertext model

Recently published formal approaches to hypertext emerge from the emphasis each puts on a special aspect of hypertext. A strong stress on linking motivates the choice of semantic networks as formal basis for hypertext in TEXTNET (Trigg/Weiser 86) or Thoth-II (Collier 87). TEXTNET employs a semantic net-work which directly connects chunks of text by semantic and rhetoric relations. The approach of Thoth-II is better suited for automatic integration of new text segments. Thoth-II is provided with a semantic network modeling the concepts within a given domain of discourse. The connection between text and conceptual knowledge is established by a string-oriented matching procedure[18]. If the handling of hypertext nodes is emphasized, object oriented approaches to hypertext are preferred (eg. Christodoulakis et al. 86, Woelk et al. 86). The planning of hypertext dialogs requires additional features like agent models—eg for the support of an extrinsic task (support of software engineering) by an agent/task driven hypertext (Garg/ Scacchi 87, Garg 87) — or constraints — the intrinsic task of planning argumentation (Smolensky et al. 87) is supported by constraints on graphical objects (on the presentation level only).

The intended hypertext model has to fit the following context:

- The first step of hypertext construction is the automatic decomposition and analysis of texts. A mapping from text units to representation structures is computed based on linguistic knowledge and background knowledge about the domain of discourse (cf. figure 4).
- Based on these representations semantic relations between text units and macrostructures of sets of text-units may be established (cf. figure 5).

The presentation of hypertext paths depends on a user formulated query[19], text plans and prototypical informational objects, which control the mapping from semantic objects to graphical objects (Thiel/ Hammwöhner 89) and thus form the elements of a graphical text presentation language (Lakin 87).

[18] Hahn 86 shows that string-oriented methods are an insufficient means for automatic text processing.

[19] The process of query formulation within a conceptual network and its support by a graphical user interface is described in Thiel/Hammwöhner 87.

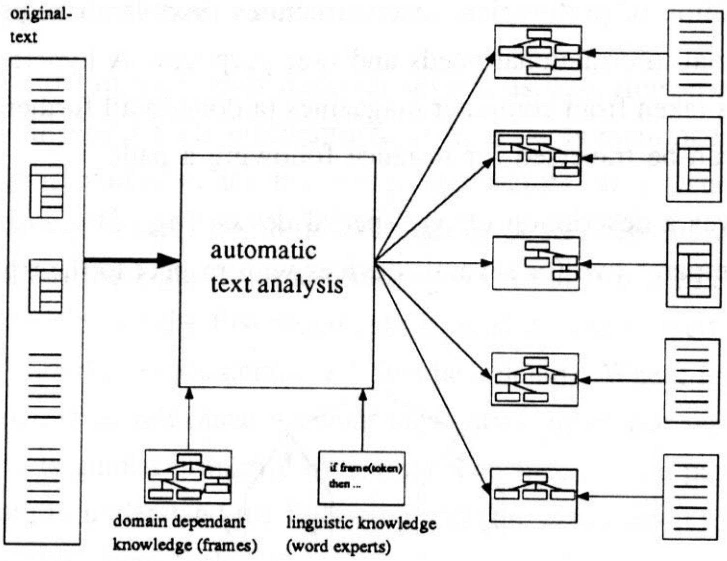

Figure 4 A text is fragmented to text units and mapped to representation structures by automatic text analysis – eg by the TOPIC system (Hahn/Reimer 86,88)

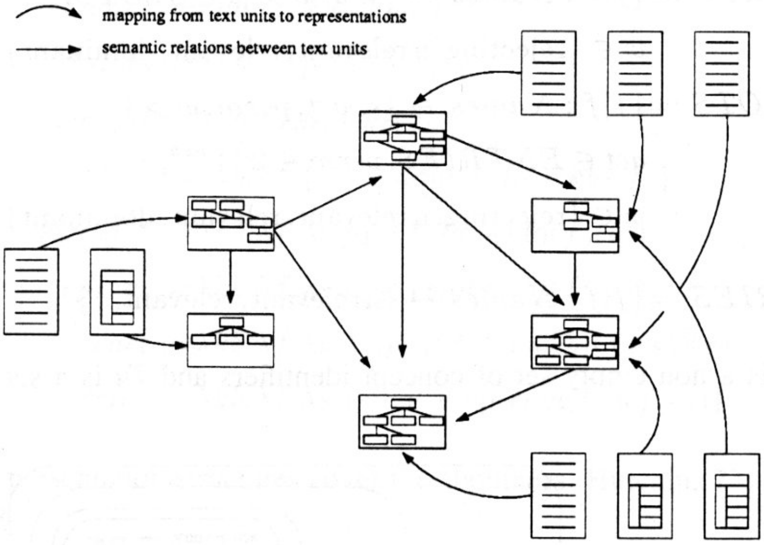
mapping from text units to representations
semantic relations between text units

Figure 5 The network structure of the hypertext depends on semantic properties of text unit representations (eg conceptual networks)

A uniform representation formalism, which can be used for conceptual knowledge, text plans and graphical structures as well, is the frame construct, which was introduced to AI by Minsky (Minsky 75). Frame like structures like case frames (Fillmore 68) or scripts (Schank/Abelson 77) are widely used in linguistics and text understanding. The relation between the constructs of frame- and object-oriented languages — the programming paradigm prevailing in computer graphics (Hollan 84) — eg. inheritance, perspectives (Bobrow/Winograd 77, Stefik/Bobrow 86) etc. is evident.

3.1 The basic representation structures

The frame formalism we will employ in our hypertext model is an extended version of FRM, the representation language of the TOPIC-system (Reimer 86, 89). The text units contained in the hypertext are mapped to sets of frames. A frame is built up by a set of slots, each of which is associated with a set of permitted and a set of actual entries. Additionally an activation weight is assigned to frames slots, and actual entries. The structure may be formalised as a cascade of partial mappings (cf figure 6):

$$HTREP := \{f \mid f : Tu \cup \{\mathrm{w}, \mathrm{q}\} \rightarrow FRAMES\} \tag{D1}$$

$$\begin{aligned} FRAMES := \{f \mid f : Names \rightarrow \{< s, w > \mid s \in SLOTS, \\ w \in \{\text{rejecting}, \text{irrelevant}, \text{relevant}, \text{dominant}\}\}\} \end{aligned} \tag{D2}$$

$$\begin{aligned} SLOTS := \{f \mid f : Names \rightarrow \{< act, perm, w > \mid \\ act \in ENTRIES, perm \in 2^{Names}, \\ w \in \{\text{rejecting}, \text{irrelevant}, \text{relevant}, \text{dominant}\}\}\} \end{aligned} \tag{D3}$$

$$ENTRIES := \{f \mid f : Names \rightarrow \{\text{irrelevant}, \text{relevant}, \}\} \tag{D4}$$

Names is a non empty set of concept identifiers and *Tu* is a set of text unit identifiers.

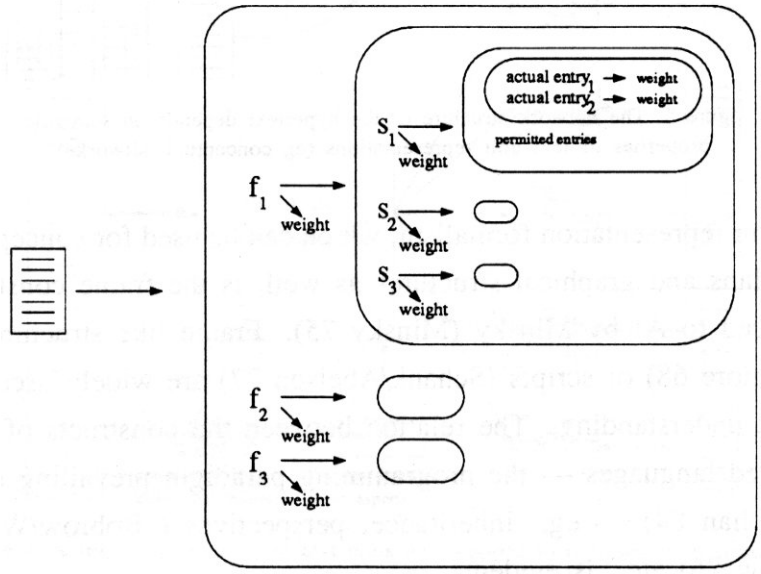

Figure 6 A hypertext representation as a cascade of mappings

Based on the cascade of mappings as described above the following functions may be defined, which allow access to:

- the frames of a text unit (D5),
- the weight of a frame (D7),
- the slots of a frame (D8),
- the weight of a slot (D9),
- the permitted entries of a slot (D10),
- the actual entries of a slot (D12),
- the weight of an entry (D11).

Elements of tuples are obtained using a projection function:

$$p_{n,i}(< x_1, x_2, .., x_i, ..x_n >) := x_i$$

$$\begin{aligned}
frames &:= \lambda kb.\{f \mid f \in dom\ kb\} && \text{(D5)}\\
fslots &:= \lambda kb.\lambda f.p_{2,1}(kb(f)) && \text{(D6)}\\
fweight &:= \lambda kb.\lambda f.p_{2,2}(kb(f)) && \text{(D7)}\\
slots &:= \lambda kb.\lambda f.\{s \mid s \in dom\ fslots(kb)(f)\} && \text{(D8)}\\
sweight &:= \lambda kb.\lambda f.\lambda s.p_{2,3}(fslots(kb)(f)(s)) && \text{(D9)}\\
eperm &:= \lambda kb.\lambda f.\lambda s.p_{2,2}(fslots(kb)(f)(s)) && \text{(D10)}\\
eact &:= \lambda kb.\lambda f.\lambda s.\lambda e.p_{2,1}(fslots(kb)(f)(s))(e) && \text{(D11)}\\
entries &:= \lambda kb.\lambda f.\lambda s.\{e \mid e \in dom\ eact(kb)(f)(s)\} && \text{(D12)}
\end{aligned}$$

These representation structures cover the aboutness (Hutchins 77) of text units as follows[20]:

- Prototype frames (frames without entries) represent the background knowledge, which is needed to understand the text unit. A set of prototypes contained in a special text unit representing the domain dependent knowledge is the basis of text analysis.

$$\begin{aligned}
prototype(kb, f) :\Leftrightarrow kb \in FRAMES \wedge f \in frames(kb) \wedge &\qquad \text{(D13)}\\
\wedge\ \forall s \in slots(kb)(f) : entries(kb)(f)(s) = \{\}&
\end{aligned}$$

- Instance frames (frames containing at least one entry) represent special knowledge as learned from the text. Every instance frame has exactly one corresponding prototype which is taken from the domain dependent knowledge.

$$\begin{aligned}
instance(kb, f) :\Leftrightarrow kb \in FRAMES \wedge f \in frames(kb) \wedge &\qquad \text{(D14)}\\
\wedge\ \exists s \in slots(kb)(f) : entries(kb)(f)(s) \neq \{\}&
\end{aligned}$$

- The activation weights indicate the salient concepts of a text unit. The most salient ones are *dominant,* followed by *relevant* ones. *Irrelevant* frames contain only background knowledge. *(Rejecting* can only occur in a query: see the D22).

The frames within a knowledge base are not just isolated objects but are interrelated by a specialization hierarchy which can be inferred from the slot-structure of frames. In this context the notion of non-terminal slots is important — these are slots the name of which is identical to a frame within the knowledge base. The permitted entries of a non-terminal slot are the subordinates of the corresponding frame.

$$non\text{-}term(kb, s) :\Leftrightarrow kb \in FRAMES \wedge s \in frames(kb) \qquad \text{(D15)}$$

[20] For a deeper understanding see Hahn/Reimer 86.

There are two specialization relations[21]:

- The *is-a* relation deals with concept specialization by adding new slots to a concept or by restricting the permitted entries of a given slot (cf figure 7).

$$\begin{aligned}
&\mathit{is\text{-}a}(kb,f,f') :\Leftrightarrow \mathit{prototype}(kb,f) \land \mathit{prototype}(kb,f')\land \\
&\quad \land \forall s' \in \mathit{slots}(kb)(f') : \exists s \in \mathit{slots}(kb)(f) : \\
&\qquad (s = s' \lor \mathit{non\text{-}term}(kb,s) \land \mathit{non\text{-}term}(kb,s') \land \mathit{e\text{-}is\text{-}a}(kb,s,s'))\land \\
&\quad \land \forall s \in \mathit{slots}(kb)(f) \cap \mathit{slots}(kb)(f') : \\
&\qquad\qquad \mathit{eperm}(kb)(f)(s) \subseteq \mathit{eperm}(kb)(f')(s)\land \\
&\quad \land \exists s \in \mathit{slots}(kb)(f) : (s \notin \mathit{slots}(kb)(f') \lor s \in \mathit{slots}(kb)(f')\land \\
&\qquad\qquad \land\, \mathit{eperm}(kb)(f)(s) \subset \mathit{eperm}(kb)(f')(s))
\end{aligned} \tag{D16}$$

- Specialization within the *inst* (instance) relation requires additional slot-entries (cf figure 7).

$$\begin{aligned}
&\mathit{inst}(kb,f,f') :\Leftrightarrow \\
&\quad (\mathit{prototype}(kb,f') \lor \mathit{instance}(kb,f')) \land \mathit{instance}(kb,f)\land \\
&\quad \land (\mathit{instance}(kb,f') \Rightarrow \\
&\qquad \Rightarrow \exists s \in \mathit{slots}(kb)(f) : \\
&\qquad\qquad \exists e \in \mathit{entries}(kb)(f)(s) : e \notin \mathit{entries}(kb)(f')(s))\land \\
&\quad \land \forall s \in \mathit{slots}(kb)(f') : \forall e' \in \mathit{entries}(kb)(f')(s) : \\
&\qquad\qquad (e' \in \mathit{entries}(kb)(f)(s) \lor \mathit{non\text{-}term}(kb,s)\land \\
&\qquad\qquad\quad \land \exists e \in \mathit{entries}(kb)(f)(s) : \mathit{e\text{-}is\text{-}a}(kb,e,e')))\land \\
&\quad \land \mathit{slots}(kb)(f) = \mathit{slots}(kb)(f')\land \\
&\quad \land \forall s \in \mathit{slots}(kb)(f) : \mathit{eperm}(kb)(f)(s) = \mathit{eperm}(kb)(f')(s)
\end{aligned} \tag{D17}$$

- The *e-is-a* relation is the transitive closure of *is-a* and *inst* (cf figure 7).

$$\begin{aligned}
&\mathit{e\text{-}is\text{-}a}(kb,f,f') :\Leftrightarrow \mathit{is\text{-}a}(kb,f,f') \lor \mathit{inst}(kb,f,f')\lor \\
&\qquad\qquad \lor \exists f'' : (\mathit{inst}(kb,f,f'') \land \mathit{is\text{-}a}(kb,f'',f'))
\end{aligned} \tag{D18}$$

[21] The formalisation given here suffices our purposes but doesn’t cover all aspects of cencept specialization in FRM as described in Reimer 86,89.

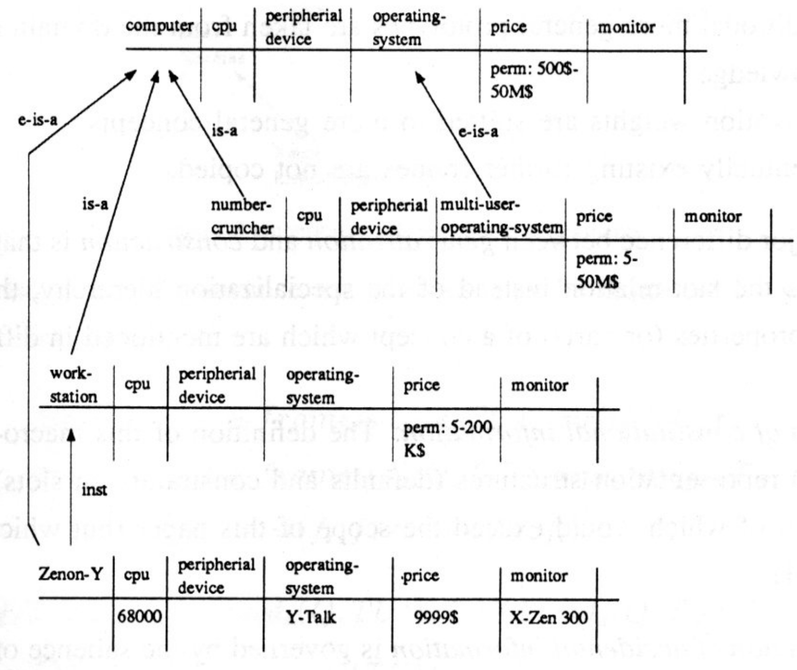

Figure 7 The is-a relation between *computer* and *number-cruncher* holds because of the restricted set of permitted entries in the *price*-slot and the specialisation of the *slot operating-system* to *multi-user-operating-system*. The inst relation between *workstation* and *Zenon-Y* holds, because of the actual slot entries.

3.2 Macro-operations in a frame-oriented hypertext model

Based on the hypertext model defined above macro-operations can be defined which allow for the derivation of abstract macro-text-units and the clustering of text units contained in a hypertext according to a user formulated query and a context oriented notion of relevance. In the following we will outline the formalization of the macro-operations as introduced above and as an example give a concise formalization of the *deletion of accidental information* (d_a).

- The *generalization* Operation g: $g \in \{ f \mid f : \text{FRAMES} \times 2^{\text{FRAMES}} \rightarrow \text{FRAMES} \}$ of a cluster of text units under consideration of domain dependent knowledge may be used as a test function for text unit clustering, because it maps improper clusters to the empty knowledge base. The units of a cluster must have salient concepts which are direct subordinates of a common prototype f_p. Each subordinate of f_p may occur only once within a cluster (to avoid uncontrolled redundancy). A (not empty) macro-text-unit a cluster of text units is mapped to fulfils the following conditions:
  a. The frames representing the salient concepts are copied to the macro-text-unit.
  b. Additional more general prototypes are taken from the domain dependent knowledge.
  c. Activation weights are shifted to more general concepts.
  d. Eventually existing further frames are not copied.
- The major difference between *generalization* and *construction* is that the latter employs the slot-relation instead of the specialization hierarchy, thus aggregating properties (or parts) of a concept which are mentioned in different text units.
- *deletion of constitutional information:* The definition of this macro-Operation requires representation structures (defaults and constraints on slots), the presentation of which would exceed the scope of this paper (but which are part of FRM).

The *deletion of accidental information* is governed by the salience of concepts within the text units on one hand and the query on the other hand. This macro-operation is a mapping from pairs of knowledge bases – queries and text units – to knowledge bases $d_a \in \{f \mid f : \mathrm{FRAMES} \times \mathrm{FRAMES} \to \mathrm{FRAMES}\}$. The resulting macro-text-units are stripped of all frames which are not relevant with respect to the query, thus irrelevant text units are mapped to the empty knowledge base. A frame f_t which represents a salient concept of the text unit TU is relevant with respect to a query Q iff there is a dominant frame f_q in Q and

- f_q has the same name as f_t or
- f_q is superordinate of ft or
- f_q has the same name as a dominant slot of f_t
- and there is no frame f_r in Q which inhibits the selection of f_t.

In this case f_t is element of d_a(Q,TU) (D19) without any difference in its structure (D20). Additionally prototypes of relevant instances are mapped to the macro-text-unit (D21). The interrelation of frames which are part of different text units – eg the closure of name identity and specialization (D23) – is based on the pre-supposition that all representations are derived from the same domain dependant knowledge and therefore contain prototypes which are common to all text units.

The inheritance of relevance assignments my be restricted by rejective assignments (D24, figure 8)

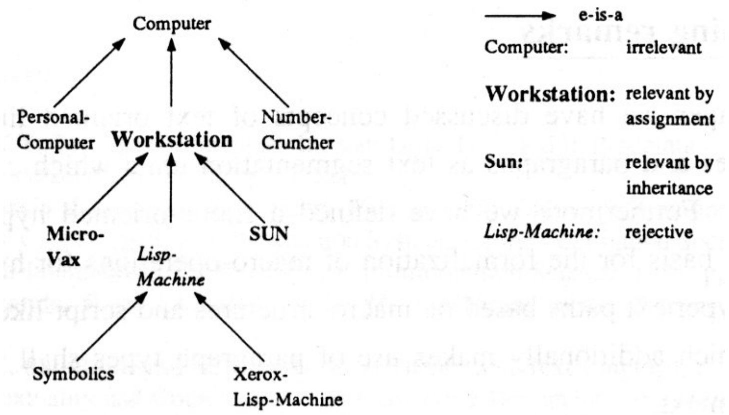

Figure 8 Inheritance of relevance assignments.

$$\forall Q, TU, f : f \in frames(d_a(Q, TU)) \Leftarrow r(Q, TU, f) \tag{D19}$$

$$\begin{aligned}\forall Q, TU, f : f \in frames(d_a(Q, TU)) \Leftarrow prototype(Tu, f) \wedge \\ \wedge \exists f_i : inst(f_i, f) \wedge r(Q, Tu, f_i)\end{aligned} \tag{D20}$$

$$\forall Q, TU, f \in frames(d_a(Q, TU)) : TU(f) = d_a(Q, TU)(f) \tag{D21}$$

$$\begin{aligned} r(Q, TU, f_t) :\Leftrightarrow f_t \in frames(TU) \wedge \\ \wedge\, (fweight(TU)(f_t) \in \{\text{dominant,relevant}\}) \wedge \\ \wedge\, \exists f_q \in frames(Q) : \\ (fweight(Q)(f_q) = \text{dominant} \wedge \\ \wedge\, (i(TU, f_t, Q, f_q) \vee f_q \in slots(TU)(f_t) \wedge \\ \wedge\, sweight(TU)(f_t)(s) = \text{dominant}) \wedge \\ \wedge\, \neg\exists f_r \in Q : b(TU, f_t, Q, f_r))\end{aligned} \tag{D22}$$

$$\begin{aligned} i(TU, f, TU', f') :\Leftrightarrow f \in frames(TU) \wedge f' \in frames(TU') \wedge \\ \wedge\, (f = f' \vee \exists f'' \in (frames(TU) \cap frames(TU')) : \\ (e\text{-}is\text{-}a(TU, f, f'') \wedge (f'' = f' \vee is\text{-}a(TU', f'', f'))))\end{aligned} \tag{D23}$$

$$\begin{aligned} b(TU, f, TU', f') :\Leftrightarrow f \in frames(TU) \wedge f' \in frames(TU') \wedge \\ \wedge\, fweight(TU', f') = \text{rejecting} \wedge i(TU, f, TU', f') \wedge \\ \wedge\, \neg\exists f'' \in frames(TU') : \\ (fweight(TU')(f'') = \text{dominant} \wedge \\ \wedge\, is\text{-}a(TU', f'', f')) \wedge i(TU, f, TU', f''))\end{aligned} \tag{D24}$$

4 Concluding remarks

In this paper we have discussed concepts of text oriented linguistics like macrostructures and paragraphs as text segmentation units which can be fruitful for hypertext. Furthermore we have defined a frame-oriented hypertext model and used it as basis for the formalization of macro-operations for hypertext. The planning of hypertext paths based on macro-structures and script-like prototypical text plans, which additionally makes use of paragraph types shall be dealt with in another context.

5 References

Ballmer, T.T. 76: Macrostructures. In: van Dijk, T. A. (ed.): Pragmatics of Language and Literature. Amsterdam, 1976, pp. 1-22.

Bates, Marcia J. 85: An Exploratory Paradigm for Online Information Retrieval. In: Brookes, B. C. (ed): Intelligent Information Systems for the Information Society. Proceedings of the Sixth International Research Forum in Information Science, 1985, pp. 91-99.

de Beaugrande, R.-A. / Dressler, W.U. 81: Einführung in die Textlinguistik, Tübingen, 1981.

Begthol, C. 86: Bibliography Classification Theory and Text Linguistics: Aboutness Analysis, Intertextuality and Cognitive Act of Classifying Documents. In: Journal of Documentation, Vol. 42, No. 2, 1986, pp. 80-113.

Belkin, N.J. / Croft, B.W. 87: Retrieval Techniques. In: Williams, Martha E. (Hg.): Annual Review of Information Science and Technology, Vol. 22, 1987, pp. 104-145.

Belkin, N. J. / Oddy, R. N. / Brooks, A. M. 82: ASK for Information Retrieval. Part I: Background and Theory. In: Journal of Documentation, Vol. 38, No. 2, 1982, pp. 61-71.

Biswas, G. / Bezdek, J. C. / Marques, M. / Subramanian, V. 87a: Knowledge-Assisted Document Retrieval: I. The Natural Language Interface. In: Journal of the American Society for Information Science, Vol. 38, No. 2, 1987, pp. 83-96.

Biswas, G. / Bezdek, J. C. / Marques, M. / Subramanian, V. 87b: Knowledge-Assisted Document Retrieval: II. The Retrieval Process. In: Journal of the American Society for Information Science, Vol. 38, No. 2, 1987, pp. 97-110.

Blair, D. C. 80: Searching Biases in Large Interactive Retrieval Systems. In: Journal of the American Society for Information Science, Vol. 31, 1980, pp 271-277.

Blair, D.C. / Maron, M.E. 85: An Evaluation of Retrieval Effectiveness for a Full Text Document Retrieval System. In: Communication of the ACM, Vol. 28, No. 3, 1985, pp. 289-299.

Bobrow, D.G. / Winograd, T. 77: An Overview of KRL-0, a Knowledge Representation Language. In: Cognitive Science, Vol. l, No. l, 1977, pp. 3-46.

Brajnik, G. / Guida, G. / Tasso, C. 87: User Modeling in Intelligent Information Retrieval. In: Information Processing & Management, Special Issue: Intelligent Information Retrieval, Vol. 13, No.4, 1987, pp. 305-320.

Burchfield, R.W. 76: A Supplement to the Oxford English Dictionary, Oxford, 1976.

Carroll, L. 39: Through the Looking-Glass. In: The Complete Works of Lewis Carroll, Glasgow, 1939, pp. 126-250.

Charney, D. 87: Comprehending Non-Linear Text: The Role of Discourse Cues and Reading Strategies. In: Proc. of the Hypertext '87 Conference, 13.-15.11.87, Univ. of NC, Chapel Hill, NC,1987, pp. 109-120.

Christodoulakis, S. / Ho, F. / Theodoridou, M. 86: The Multimedia Object Presentation Manager of Minos: A Symmetry Approach. In: Sigmod Record, Vol. 15, No. 2, 1986, pp. 295-310.

Collier, G.H. 87: Thoth-II Hypertext with Explicit Semantics. In: Hypertext '87 Papers, Chapel Hill, NC, University of North Carolina, 1987, pp. 269-289.

Conklin, J. 87: A Survey on Hypertext, MCC Technical Report No STP-356-86, Rev. l, 1987.

Croft, W.B. / Thompson, R.H. 87: I^{3}R: A New Approach to the Design of Document Retrieval Systems. In: Journal of the American Society for Information Science, Vol. 38, No. 6, 1987, pp. 389-404.

Daneš, F. 78: Zur linguistischen Analyse der Textstruktur. In: Dressler, Wolfgang (ed): Textlinguistik, Darmstadt, 1978, pp. 184-192.

Defude, B. / Chiamarella, Y. 87: A Prototype of an Intelligent System for Information Retrieval: IOTA. In: Information Processing & Management, Vol. 13, No. 4, 1987, pp. 284-303.

van Dijk, T. 79: Relevance Assignment in Discourse Comprehension. In: Discourse Processes, Vol. 2, No.2, 1979, pp. 113-126.

van Dijk, T. 80a: Macrostructures, Hillsdale, 1980.

van Dijk, T. 80b: Textwissenschaft, München, 1980.

Fillmore, CJ. 68 The Case for Case. In: Bach, E. / Harms, R.T. (eds): Universals in Linguistic Theory. New York, 1968, pp. 1-88.

Frisse, M.E. 87: Searching for Information in a Hypertext Medical Handbook. In: Proc. of the Hypertext '87 Conference, 13.-15.11.87, Univ. of NC, Chapel Hill, NC,1987, pp. 57-66.

Frisse, M.E. 88: From Text to Hypertext. In: Byte, Vol. 13, No. 10, 1988, pp. 247-253.

Fum, D. / Guida, G./ Tasso, C. 82: Forward and Backward Reasoning in Automatic Abstracting. In: Proc. 9th Int. Joint. Conf. on Artificial Intelligence, 1985, pp. 840-844.

Garcia-Berrio, A. / Mayordomo, T.A. 88: Compositional Structure: Macrostructure. In: Petöfi, JanoS (ed): Text and Discourse Constitution. Berlin et al., 1988, pp 170-211.

Garg, P.K. 87: Abstraction Mechanisms in Hypertext. In: Hypertext '87 Papers, Chapel Hill, NC, University of North Carolina, 1987, pp. 375-395

Garg, P.K. / Scacchi, W. 87: On Designing Intelligent Hypertext Systems for Information Management in Software Engineering. In: Hypertext '87 Papers, Chapel Hill, NC, University of North Carolina, 1987, pp. 409-432

Garnes, Sara 87: Paragraph Perception by Seven Groups of Readers. Ohio State University Working Papers in Linguistics, 1987, Vol. 35, pp. 132-141.

Giora, R. 83: Functional Paragraph Perspective. In: Petöfi, Janös S./ Sözer, Emel (ed):Micro and Macro Connexity of Texts. Hamburg, 1983, pp. 153-182.

Grice, H.P. 75: Logic and Conversation. In: Cole, R. / Morgan, J. L. (ed): Syntax and Semantics, Vol. 3, Speech Acts, New York et al., 1975, pp. 41-58. **Große, E.U.** 74: Texttypen, Stuttgart et al. 1974.

Gülich, E. / Raible, W. 77: Linguistische Textmodelle. Basel etc. 1977.

Hahn, U. 86: Methoden der Volltextverarbeitung in Informationssystemen. In: Kuhlen, Rainer (Hg.): Informationslinguistik. Tubingen, 1986, pp. 195-216.

Hahn, U. / Reimer, U. 86: TOPIC-Essentials. In: COLING-86. Proceedings of the llth International Conference on Computational Linguistics, 1986, pp. 497-503.

Hahn, U. / Reimer, U. 88: Knowledge-Based Text Analysis in Office Environments: The Text Condensation System TOPIC. In: Lamersdorf, W. (ed): Office Knowledge: Represen-tation, Management and Utilization. Amsterdam, 1988, pp. 197-215.

Hammwöhner, R. / Thiel, U. 87: Content Oriented Relations between Text Units — A Structural Model for Hypertexts. In: Hypertext '87 Papers, Chapel Hill, NC, University of North Carolina, 1987, pp. 155-174.

Hobbs, J.R. 83: Why is Discourse Coherent? In: Neubauer, Fritz (ed): Coherence in Natural Language Texts. Papiere zur Textlinguistik Band 38, Hamburg, 1983, pp. 29-70.

Hollan, J.D. 84: Intelligent Object-Based Graphical Interfaces. In: Salvendi, G. (ed): Human—Computer Interaction. Amsterdam, 1984, pp. 293-296.

Hutchins, W.J. 77: On the Problem of 'Aboutness' in Document Analysis. In: Journal of Informatics, Vol. l, No. l, 1977, pp. 17-35.

Hutchins, J. 87: Summarization : Some Problems and Methods. In: K. P. Jones (ed.): Informatics 9, Proc. of a Conference ... King's College, Cambridge, 26-27 March 1987, London, 1987, pp. 151-173.

Jones, W.P. 87: How Do We Distinguish the Hyper from the Hype in Non-linear Text? In: Bullinger, HJ./ Shackel, B./. Kornwachs, K. (eds): Human-Computer Interaction — INTERACT 87, Proceedings of the Second IFIP Conference on Human-Computer Interaction, Univ. Stuttgart, ERG, 1.-4. Sept. 1987.

Kieras, D.E. 82: A Model of Reader Strategy for Abstracting Main Ideas from Simple Technical Prose. In: Text, Vol. 2, No. 1-3, 1982, pp. 47-81.

Koen, F. / Becker, A./ Young, R. 69: The Psychological Reality of the Paragraph. Journal of Verbal Learning and Verbal Behaviour, Vol. 8, No. l, 1969, pp. 49-53.

Lakin, F. 87: Visual Grammars for Visual Languages. In: AAAI87 — Proceedings 6th Nat. Conf. on Art. Int., Vol. II, Los Altos, 1987, pp. 683-688

Larson, R.P. 88: Hypertext and Information Retrieval: Towards the Next Generation of Information Systems. In: Information & Technology, Proc. of the 51st Meeting of the American Society for Information Science, Atlanta, Georgia, 1988, pp. 195-199.

Longacre, R. E. 74: Sentence Structure as a Statement Calculus. In: Brend, Ruth M. (ed): Advances in Tagmemics. Amsterdam et al., 1974, pp. 251-283.

Longacre, R. E. 76: Discourse. In: Brend, Ruth M./Pike, Kenneth L. (ed): Tagmemics — Aspects of the Field. Paris, 1976, pp. 1-44.

Longacre, R. E. 79: The Paragraph as a Grammatical Unit. In: Givon, T. (ed): Syntax and Semantics 12, New York, 1979.

Mann, W.C. / Thompson, S.A. 88: Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. In: Text, Vol. 8, No. 3, 1988, pp. 243-281.

Minsky, M. 75: A Framework for Representing Knowledge. In: Winston, P. (ed.): The Psychology of Computer Vision, New York: McGraw Hill, 1975, pp. 211-277.

Petöfi, J.S. 79: Die Struktur der TeSWeST. Aspekte der pragmatisch-semantischen Interpretation von objektsprachlichen Texten. In: Neubauer, Fritz (ed): Coherence in Natural Language Texts. Hamburg, 1979.

Phillips, M. 85: Aspects of Text Structure — An Investigation of the Lexical Organization of Text. Amsterdam et al., 1985.

Pike, K.L. / Pike, E.G. 77: Grammatical Analysis. Dallas, 1977.

Rau, L.F. 87a: Information Retrieval from Never-ending Stories. In: AAAI87 — Proc. 6th Nat. Conf. on Art. Int., Vol. I, Los Altos; Morgan Kaufman, 1987, pp 317-321

Rau, L.F. 87b: Spontaneous Retrieval in a Conceptual Information System. In: Proc. of the lOth Int. Joint Conf. on AI, Milan, 1987, pp. 155-162.

Reimer, U. **86:** A System-Controlled Multi-Type Specialization Hierarchy. In: Kerschberg, L. (ed.): Expert Database Systems. Proceedings of the Ist International Workshop, Menlo Park/CA: Benjamin/Cummings, 1986, pp. 173-187.

Reimer, U. **89:** FRM: Ein Frame-Repräsentationsmodell und seine formale Semantik. Berlin, Heidelberg, 1989.

Reimer, U. / Hahn, U. 88: Text Condensation as Knowledge Base Abstraction. In: Proceedings — The Fourth IEEE Conference on Artificial Intelligence Application, San Diego, California, Washington, D.C.: Comp. Soc. of the IEEE, 1988.

Remde, J.R. / Gomez, L.M./ Landauer, T.K. 87: SuperBook an Automatic Tool for Information Exploration — Hypertext? In: Hypertext '87 Papers, Chapel Hill, NC, University of North Carolina, 1987, pp. 175-188.

Robertson, S.E. 80: Some Recent Theories and Models in Information Retrieval. In: Harbo, O. / Kajberg, C. (Hg.): Theory and Applications of Information Research, London, 1980, pp. 131-136.

Schank, K. / Abelson, R. 77: Scripts, Plans, Goals, and Understanding. Hillsdale, 1977.

Rumelhart, D.E. 75: Notes on a Schema for Stories. In: Bobrow, D.G./ Collins, A. (eds): Representation and Understanding: Studies in Cognitive Science. New York, 1975, pp. 211-236.

Simmons, R.F. 87: A Text Knowledge Base from the AI Handbook. In: Information Processing & Management, Vol. 13, No. 4, 1987, pp. 321-339.

Smith, L.C. 87: Artificial Intelligence and Information Retrieval. In: Williams, Martha E. (Hg.): Annual Review of Information Science and Technology, Vol. 22, 1987, pp. 41-77.

Smolensky, P. / Bell, B. / Fox, B. / King, R. / Lewis, C. 87: Constraint-Based Hypertext for Argumentation. In: Hypertext '87 Papers, Chapel Hill, NC, University of North Carolina, 1987, pp. 215-245.

Sonnenberger, G. 88: Flexible Generierung von natürlichsprachigen Abstracts aus Textrepräsentationsstrukturen. In: Trost, H. (ed): 4. Österreichische Artificial Intelligence Tagung, Wiener Workshop — Wissensbasierte Sprachverarbeitung, Berlin etc, 1988, pp. 72-82.

Stark, H.A. 88: What do Paragraph Markings Do? In: Discourse Processes, No. 11, Vol. 3, 1988, pp. 275-303.

Stefik, M. / Bobrow, D.G. 86: Object Oriented Programming: Themes and Variations. In: The AI-Magazine, Vol. 6, No. 4, 1986, pp. 40-62.

Stibic, V. 85: Printed Versus Displayed Information. In: Nachrichten für Dokumentation , Vol. 36, No. 4/5, 1985, pp. 172-178.

Swanson, D.R. 77: Information Retrieval äs a Trial-and-Error Process. In: Library Quaterly, Vol. 47, No. 2, 1977, pp. 128-148.

Tenopir, C., 85: Full-Text Database Retrieval Performance. In: Online Review, Vol. 9, No. 2, 1985, pp. 149-164.

Thiel, U. / Hammwöhner, R. 86: Graphical Interaction with a Full-Text Oriented Information System: The Retrieval Component of the End User Interface TOPOGRAPHIC.In: Proc. of the 2nd Int. Conf. on the Application of Micro-Computers in Information, Documentation and Libraries. Amsterdam etc., 1986.

Thiel, U. / Hammwöhner, R. 87: Informational Zooming: An Interaction Model for the Graphical Access to Text Knowledge Bases, in: Yu,C. T. / van Rijsbergen, C. J. (eds): Proc. lOth Annual Int. ACMSIGIR Conf. on Research & Development in Information Retrieval, New Orleans, Louisiana, USA, 1987, pp. 45-56.

Thiel, U. / Hammwöhner, R. 89: Interaktion mit Textwissensbasen — Ein objektorientierter Ansatz. In: Tagungsband der Jahrestagung der Gesellschaft für Informatik 1989, to appear.

Tiamiyu, M. / Ajiferuke, I.Y. 88: A Total Relevance and Document Interaction Effects Model for the Evaluation of Information Retrieval Processes. In: Information Processing & Management, Vol. 24, No. 4, 1988, pp. 391-404.

Trigg, R.H. / Weiser, M. 86: TEXTNET: A Network-Based Approach to Text Handling. In:ACM Transactions on Office Information Systems, Vol. 4., No. l, 1986, pp 1-23.

Woelk, D. / Kim, W. / Luther, W. 86: A Object Oriented Approach to Multimedia Databases. In SIGMOD Record, Vol. 15, No.2, 1986, pp. 311-325.

Zimmermann, K. 78: Erkundungen zur Texttypologie, Tübingen, 1978.