

TEXTUAL EXPERTISE IN WORD EXPERTS:  
AN APPROACH TO TEXT PARSING BASED ON TOPIC/COMMENT MONITORING \*

Udo Hahn

Universitaet Konstanz  
Informationswissenschaft  
Projekt TOPIC  
Postfach 5560  
D-7750 Konstanz 1, West Germany

ABSTRACT

In this paper prototype versions of two word experts for text analysis are dealt with which demonstrate that word experts are a feasible tool for parsing texts on the level of text cohesion as well as text coherence. The analysis is based on two major knowledge sources: context information is modelled in terms of a frame knowledge base, while the co-text keeps record of the linear sequencing of text analysis. The result of text parsing consists of a text graph reflecting the thematic organization of topics in a text.

1. Word Experts as a Text Parsing Device

This paper outlines an operational representation of the notion of text cohesion and text coherence based on a collection of word experts as central procedural components of a distributed lexical grammar.

By text cohesion, we refer to the micro level of textuality as provided, e.g. by reference, substitution, ellipsis, conjunction and lexical cohesion (cf. HALLIDAY/HASAN 1976), whereas text coherence relates to the macro level of textuality as induced, e.g. by patterns of semantic recurrence of topics (thematic progression) of a text (cf. DANES 1974). On a deeper level of propositional analysis of texts further types of semantic development of a text can be examined, e.g. coherence relations, such as contrast, generalization, explanation (cf. HOBBS 1979, HOBBS 1982, DIJK 1980a), basic modes of topic development, such as expansion, shift, or splitting (cf. GRIMES 1978), and operations on different levels of textual macro-structures (DIJK 1980a) or schematized superstructures (DIJK 1980b).

The identification of cohesive parts of a text is needed to determine the continuous development and increment of information with regard to single thematic foci, i.e. topics of the text. As we have topic elaborations, shifts, breaks, etc. in texts the extension of topics has to be delimited exactly and different topics have to be related properly. The identification of coherent parts of a text serves this purpose, in that the determination of the coherence relations mentioned above

contributes to the delimitation of topics and their organization in terms of text grammatical well-formedness considerations. Text graphs are used as the resulting structure of text parsing and serve to represent corresponding relations holding between different topics.

The word experts outlined below are part of a genuine text-based parsing formalism incorporating a linguistical level in terms of a distributed text grammar and a computational level in terms of a corresponding text parser (HAHN/REIMER 1983; for an account of the original conception of word expert parsing, cf. SMALL/RIEGER 1982). This paper is intended to provide an empirical assessment of word experts for the purpose of text parsing. We thus arrive at a predominantly functional description of this parsing device neglecting to a large extent its procedural aspects.

The word expert parser is currently being implemented as a major system component of TOPIC, a knowledge-based text analysis system which is intended to provide text summarization (abstracting) facilities on variable layers of informational specificity for German language texts (each approx. 2000-4000 words) dealing with information technology. Word expert construction and modification is supported by a word expert editor using a special word expert representation language fragments of which are introduced in this paper (for a more detailed account, cf. HAHN/REIMER 1983, HAHN 1984). Word experts are executed by interpretation of their representation language description. TOPIC's word expert system and its editor are written in the C programming language and are running under UNIX.

2. Some General Remarks about Word Expert Structure and the Knowledge Sources Available for Text Parsing

A word expert is a procedural agent incorporating linguistic and world knowledge about a particular word. This knowledge is represented declaratively in terms of a decision net whose nodes are constructed of various conditions. Word experts communicate among each other as well as with other system components in order to elaborate a word's meaning (reading).

The conditions at least are tested for two kinds of knowledge sources, the context and the co-text of the corresponding word.

\* Work reported in this paper is supported by BMFT/GID under grant no. PT 200.08.

Context is a frame knowledge base which contains the conceptual world knowledge relevant for the texts being processed. Simple conditions to be tested in that knowledge base are:

```

ACTIVE ( f ) : <==>
  f is an active frame
EISA ( f , f' ) : <==>
  frame f is subordinate or instance of
  frame f'
HAS_SLOT ( f , s ) : <==>
  frame f has slot s associated to it
HAS_SVAL ( f , s , v ) : <==>
  slot s of frame f has been assigned the
  slot value v
SVAL_RANGE ( str , s , f ) : <==>
  string str is a permitted slot value with
  respect to slot s of frame f

```

Co-text is a data repository which keeps record of the sequential course of the text analysis actually going on - this linear type of information is completely lost in the context, although it is badly needed for various sorts of textual cohesion and coherence phenomena. As co-text necessarily reflects basic properties of the frame representation structures underlying the context, some conditions to be tested in the co-text also take certain aspects of context knowledge into account:

```

BEFORE ( exp , str1 , str2 ) : <==>
  str1 occurs maximally exp many trans-
  actions before str2 in the co-text
AFTER ( exp , str1 , str2 ) : <==>
  str1 occurs maximally exp many trans-
  actions after str2 in the co-text
IN_PHRASE ( str1 , str2 ) : <==>
  str1 occurs in the same sentence as str2
EQUAL ( str1 , str2 ) : <==>
  str1 equals str2
FACT ( f ) : <==>
  frame f was affected by an activation op-
  eration in the knowledge base
SACT ( f , s ) : <==>
  slot s of frame f was affected by an ac-
  tivation operation in the knowledge base
SVAL ( f , s , v ) : <==>
  slot s of frame f was affected by the as-
  signment of a slot value v in the know-
  ledge base
SAME_TRANSACTION ( f , f' ) : <==>
  frame f and frame f' are part of the same
  transaction with respect to a single text
  token, i.e. the set of all operations on
  the frame knowledge base which are car-
  ried out due to the readings generated by
  the word experts which have been put into
  operation with respect to this token

```

From the above atomic predicates more complex conditions can be generated using common logical operators (AND, OR, NOT). These expressions underlie an implicit existential quantification, unless specified otherwise.

During the operation of a word expert the variables of each condition have to be bound in order to work out a truth value. In App.A and App.B

underlining of variables indicates that they have already been bound, i.e. the evaluation of the condition in which a variable occurs takes the value already assigned, otherwise a value assignment is made which satisfies the condition being tested.

Items stored in the co-text are in the format

TOKEN	actual form of text word
TYPE	normalized form of text word after morpho-         logical reduction or decomposition proce-         dures have operated on it
ANNOT	annotation indicating whether TYPE is iden-         tified as
	FRAME a frame name
	WEXP a word expert name
	STOP a stop word or
	NUM a numerical string
	NIL an unknown text word
	or TYPE consists of parameters
	frame . slot . sval
	which are affected by a special type of op-         eration executed in the frame knowledge         base which is alternatively denoted by
	FACT frame activation
	SACT slot activation
	SVAL slot value assignment

### 3. Two Word Experts for Text Parsing

We now turn to an operational representation of the notions introduced in sec.1. The discussion will be limited to well-known cases of textual cohesion and coherence as illustrated by the following text segment:

- [1] In seiner Grundversion ist der Mikrocomputer mit einem Z-80 und 48 KByte RAM ausgeruestet und laeuft unter CP/M. An Peripherie werden Tastatur, Bildschirm und ein Tintenspritzdrucker bereitgestellt. Schliesslich verfuegt das System ueber 2 Programmiersprachen: Basic wird von SystemSoft geliefert und der Pascal-Compiler kommt von PascWare.

[The basic version of the micro is supplied with a Z-80, 48 kbyte RAM and runs under CP/M. Peripheral devices provided include a keyboard, a CRT display and an ink jet printer. Finally, the system makes available 2 programming languages: Basic is supplied by SystemSoft while PascWare furnished the Pascal compiler.]

First, in sec.3.1 we will examine textual cohesion phenomena illustrated by special cases of lexical cohesion, namely the tendency of terms to share the same lexical environment (collocation of terms) and the occurrence of 'general nouns' referring to more specific terms (cf. HALLIDAY/HASAN 1976). Then, in sec.3.2 our discussion will be centered around various modes of thematic progression in texts, such as linear thematization of rhemes (cf. DANES 1974) which is often used to establish text coherence (for a similar approach to combine the topic/comment analysis of texts and knowledge representation based on the frame model,

cf. CRITZ 1982; computational analysis of textual coherence is also provided by HOBBS 1979, 1982 applying a logical representation model).

Word experts capable of handling corresponding textual phenomena are given in App.A and App.B. However, only simplified versions of word experts (prototypes) can be supplied restricting their scope to the recognition of the text structures under examination. The representation of the textual analysis also lacks completeness skipping a lot of intermediary steps concerning the operation of other (e.g. phrasal) types of word experts (for more details, cf. HAHN 1984).

### 3.1 A Word Expert for Text Cohesion

We now illustrate the operation of the word expert designed to handle special cases of text cohesion (App.A) as indicated by text segment [1].

Suppose, the analysis of the text has been carried out covering the first 9 text words of [1] as indicated by the entries in co-text:

NO.	TOKEN	TYPE	ANNOT
{01}	In	in	STOP
{02}	seiner	sein	STOP
{03}	Grundversion	-	NIL
{04}	ist	ist	STOP
{05}	der	der	STOP
{06}	Mikrocomputer	Mikrocomputer	FRAME
{07}	mit	mit	STOP
{08}	einem	ein	STOP
{09}	Z-80	Z-80	FRAME

The word expert given in App.A starts running whenever a frame name occurs in the text. Starting at the occurrence of frame "Mikrocomputer" indicated by {06} no reading is worked out. At {09} the expert's input variable 'frame' is bound to "Z-80" as it starts again. A test in the knowledge base indicates that "Z-80" is an active frame (by default operation). Proceeding backwards from the current entry in co-text the evaluation of nodes #10 and #11 yields TRUE, since pronoun list contains an element "ein" a morphological variant of which occurs immediately before frame (Z-80) within the same sentence. In addition, we set frame' to "Mikrocomputer" (micro computer) as it is next before frame (with proximity left unconstrained due to 'any') in correspondence with {06}, and it is an active frame, too. The evaluation of node #12, finally, produces FALSE, since frame' (Mikrocomputer) is not a subordinate or instance of frame (Z-80) - actually, "Z-80" is an instance of "Mikroprocessor" (micro processor). Following the FALSE arc of #12 leads to expression #2 which evaluates to FALSE, as frame' (Mikrocomputer) is a frame which roughly consists of the following set of slots (given by indentation)

Mikrocomputer	micro computer
Mikroprocessor	micro processor
Peripherie	peripheral devices
Hauptspeicher	main memory
Programmiersprache	programming language
Systemsoftware	system software

Following the FALSE arc of #2, #3 also evaluates to FALSE as according to the current state of analysis context contains no information indicating that frame' (Mikrocomputer) has a slot' to which has been assigned any slot value (in addition, "Z-80" is not used as a default slot value of any of the slots supplied above). Turning now to the evaluation of #4 slot' has to be identified which must be a slot of frame' (Mikrocomputer) and frame (Z-80) must be within the value range of permitted slot values for slot' of frame'. Trying "Mikroprozessor" for slot' succeeds, as "Z-80" is an instance of "Mikroprozessor" and thus (due to model-dependent semantic integrity constraints inherent to the underlying frame data model [REIMER/HAHN 1983]) it is a permitted slot value with respect to slot' (Mikroprozessor) which in turn is a slot of frame' (Mikrocomputer). Thus, the interpretation slot' as "Mikroprozessor" holds. The execution of word experts terminates if a reading has been generated. Readings are labels of leaf nodes of word experts, so following the TRUE arc of #4 the reading SVAL ASSIGN ( Mikrocomputer , Mikroprozessor , Z-80 ) is reached. SVAL ASSIGN\* is a command issued to the frame knowledge base (as is done with every reading referring to cohesion properties of texts) which leads to the assignment of the slot value "Z-80" to the slot "Mikroprozessor" of the frame "Mikrocomputer". This operation also gets recorded in co-text (SVAL). Therefore, entry {09} get augmented:

NO.	TOKEN	TYPE	ANNOT
{09}	Z-80	Z-80	FRAME SVAL

The next steps of the analysis are skipped, until a second basic type of text cohesion can be examined with regard to {34}:

NO.	TOKEN	TYPE	ANNOT
{11}	48	48	NUM
{12}	KByte	KByte	WEKP
{13}	RAM	RAM	FRAME
	-	RAM-1.Groesse.48 KByte	SVAL
{18}	CP/M	Mikrocomputer.Hauptspeicher.RAM-1	FRAME
	-	CP/M	SVAL
{19}	.	Mikrocomputer.Betriebssystem.CP/M	WEKP
{21}	Peripherie	Peripherie	FRAME
	-	Mikrocomputer.Peripherie	SACT
{23}	Tastatur	Tastatur	FRAME
	-	Mikrocomputer.Peripherie.Tastatur	SVAL
{25}	Bildschirm	Bildschirm	FRAME
	-	Mikrocomputer.Peripherie.Bildschirm	SVAL
{28}	Tintenspritzdrucker	Tintenspritzdrucker	FRAME
	-	Mikrocomputer.Peripherie.Tintenspritzdrucker	SVAL
{30}	.	.	WEKP
{33}	das	das	STOP
{34}	System	System	FRAME

At {34} the word expert dealing with text cohesion phenomena again starts running. Its input variable 'frame' is set to "System" (system). With respect to #10 the evaluation of BEFORE yields a positive result, since "das" which is an element of pronoun list occurs immediately before frame. As the

\* SWEIGHT\_INC (f, s) which is also provided in App.A says that the activation weight of slot s of frame f gets incremented.

IN\_PHRASE predicate also evaluates to TRUE, the whole expression #10 turns out to be TRUE. Proceeding backwards to the next frame which is active in the frame knowledge base search stops at position {28}. When more than a single frame within the same transaction may be referred to by word experts the following reference convention is applied:

- [2i] if ANNOT = FRAME and an annotation of type FACT exists examine the frame corresponding to FACT
- [2ii] if ANNOT = FRAME or ANNOT = WEXP and annotations of type SACT or SVAL exist examine f as frame, s as slot, and v as slot value, resp. according to the order of parameters f . s . v

In these cases reference of word experts to the frame corresponding to the annotation FRAME would cause the provision of insufficient or even false structural information about the context of the current lexical item, although more significant information actually is available in the knowledge sources. In the word expert considered, frame' is set to "Mikrocomputer" according to [2ii]. Following the TRUE arc of #11 expression #12 states that frame' (Mikrocomputer) must be a subordinate or instance of frame (System) which also holds TRUE. Thus, one gets the reading SHIFT ( System , Mikrocomputer ) which says that the activation weight of frame (System) has to be decremented (thus neutralizing the default activation), while the activation weight of frame' (Mikrocomputer) gets incremented instead. Based on this re-assignment of activation weights the system is protected against invalid activation states, since "Mikrocomputer" is referred to by "System" due to stylistic reasons only and no indication is available that a real topical change in the text is implied, e.g. some generalization with respect to the whole class of micro computers. We thus have an augmented entry for {34} in co-text together with the result of processing the remainder of [1]:

NO.	TOKEN	TYPE	ANNOT
{34}	System	System	FRAME
	-	Mikrocomputer	
{36}	2	2	FACT
{37}	Programmiersprachen	Programmiersprache	FRAME
	-	Mikrocomputer.Programmiersprache.	SACT
{39}	Basic	Basic	FRAME
	-	Mikrocomputer.Programmiersprache.Basic	SVAL
{42}	SystemSoft	SystemSoft	FRAME
	-	Basic.Hersteller.SystemSoft	SVAL
{46}	Pascal-Compiler	Pascal-Compiler	FRAME
	-	Mikrocomputer.SystemSoftware.Pascal-Compiler	SVAL
	-	Pascal	FRAME
	-	Mikrocomputer.Programmiersprache.Pascal	SVAL
{49}	PascWare	PascWare	FRAME
	-	Pascal-Compiler.Hersteller.PascWare	SVAL
	-	Pascal.Hersteller.PascWare	SVAL

While expressions #1-#4 of App.A handle the usual kind of lexical cohesion sequencing in German a variant form of lexical cohesion is provided for by #5-#8 with reverse order of sequencing ("... die Tastatur fuer den Mikrorechner ..." or "... die Tastatur des Mikros ..."). From this outline one gets a slight impression of the text parsing capabilities inherent to word experts on the level of text cohesion as parsing is performed irrespective of sentence boundaries on a primarily semantic level of text processing in a non-expensive way

(partial parsing). With respect to other kinds of cohesive phenomena in texts, e.g. pronominal anaphora, conjunction, deixis, word experts are available similar in structure, but adapted to identify corresponding phenomena.

### 3.2 A Word Expert for Text Coherence

We now examine the generation of a second type of reading, so-called coherence readings, concerning the structural organization of cohesive parts of a text. Unlike cohesion readings, coherence readings of that type are not issued to the frame knowledge base to instantiate various operations, but are passed over to a data repository in which coherence indicators of different sorts are collected continuously. A device operating on these coherence indicators computes text structure patterns in terms of a text graph which is the final result of text parsing in TOPIC.

A text graph constructed that way is composed of a small set of basic coherence relations. We only mention here the application of further relations due to other types of linguistic coherence readings (cf. HAHN 1984) as well as coherence readings from computation procedures based exclusively on configuration data from the frame knowledge base (HAHN/REIMER 1984). One common type of coherence relations is accounted for in the remainder of section which provides for a structural representation of texts which is already well-known following DANES' 1974 distinction among various patterns of thematic progression:

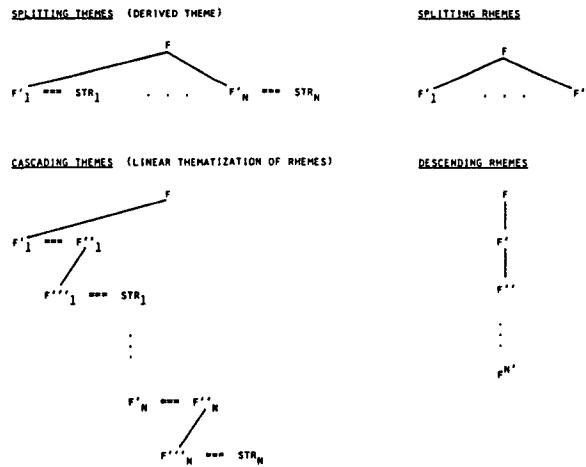


Fig.1: Graphical Interpretation of Patterns of Thematic Progression in Texts

The meaning of the coherence readings provided in App.B with respect to the construction of the text graph is stated below:

SPLITTING\_RHEMES ( f , f' )  
frame f is alpha ancestor to f'  
DESCENDING\_RHEMES ( f , f' , f'' )  
frame f is alpha ancestor to f' &  
frame f' is alpha ancestor to f''

```

CONSTANT_THEME ( f , str )
    frame f is beta ancestor string str
SPLITTING_THEMES ( f , f' , str )
    frame f is alpha ancestor to f' &
        frame f' is beta ancestor to string str
CASCADING_THEMES ( f , f' , f'' , f''' , str )
    frame f is alpha ancestor f' &
        frame f' is beta ancestor to f'' &
            frame f'' is alpha ancestor to f''' &
                frame f''' is beta ancestor to string str
SEPARATOR ( f )
    frame f is alpha ancestor to a separator
        symbol

```

We now illustrate the operation of the word expert designed to handle special cases of text coherence (App.B) as indicated by text segment [1]. It gets started whenever a frame name has been identified in the text. Suppose, we have frame set to "Mikrocomputer" with respect to {06}. Since #1 fails (there is no other frame' available within transaction {06}), evaluating #2 leads to the assignment of "Mikrocomputer" to frame' (with respect to {09}), since according to convention [21] and to the entries of co-text frame' (Mikrocomputer/{09}) occurs after frame and is immediately adjacent to frame (Mikrocomputer/06}); in addition, both, frame as well as frame', belong to different transactions. Thus, #2 is evaluated TRUE. Obviously, #3 also holds TRUE, whereas #4 evaluates to FALSE, since frame' is annotated by SVAL according to the co-text instead of SACT, as is required by #4. Note that only the same transaction (if #1 holds TRUE) or the next transaction (if #2 holds TRUE) is examined for appropriate occurrences of SACTs or SVALs. With respect to #5 the SVAL annotation covers the following parameters in {09}: frame' (Mikrocomputer), slot' (Mikroprozessor) and sval' (Z-80). Proceeding to the next state of the word expert (#6) we have frame (Mikrocomputer) but no SVAL or SACT annotation with respect to {06}. Thus, #6 necessarily gets FALSE, so that, finally, the reading SPLITTING\_THEMES (Mikrocomputer , Mikroprozessor , Z-80 ) is generated.

A second example of the generation of a coherence reading starts setting frame to "RAM-1" at position {13} in the co-text. Evaluating #1 leads to the assignment of "Mikrocomputer" to frame', since two frames are available within the same transaction. Both frames being different from each other one has to follow the FALSE arc of #3. Similar to the case above, both transaction elements in {13} are annotated by SVAL, such that #7 as well as #9 are evaluated FALSE, thus reaching #11. Since frame (RAM-1) has got no slot to which has been assigned frame' (Mikrocomputer), #11 evaluates to FALSE. With respect to #13 we have frame' (Mikrocomputer) whose slot' (Hauptspeicher) has been assigned a slot value which equals frame (RAM-1). At #14, finally, slot (Groesse) and sval (48 KByte) are determined with respect to frame (RAM-1). The coherence reading worked out is stated as CASCADING\_THEMES ( Mikrocomputer , Hauptspeicher , RAM-1 , Groesse , 48 KByte ).

Completing the coherence analysis of text segment [1] at last yields the final expansion of

co-text (note that both word experts described operate in parallel, as they are activated by the same starting criterion):

	READING	PARAMETERS
9)	SPLITTING_THEMES	Mikrocomputer.Mikroprozessor.Z-80
13)	SPLITTING_THEMES	Mikrocomputer.Hauptspeicher.RAM-1
	CASCADING_THEMES	Mikrocomputer.Hauptspeicher.RAM-1.Groesse.48 KByte
18)	SPLITTING_THEMES	Mikrocomputer.Betriebssystem.CP/M
21)	SPLITTING_RHEMES	Mikrocomputer.Peripherie
23)	SPLITTING_THEMES	Mikrocomputer.Peripherie.Tastatur
25)	SPLITTING_THEMES	Mikrocomputer.Peripherie.Bildschirm
28)	SPLITTING_THEMES	Mikrocomputer.Peripherie.Tintenstrudrucker
34)	SEPARATOR	Mikrocomputer
37)	SPLITTING_RHEMES	Mikrocomputer.Programmiersprache
40)	SPLITTING_THEMES	Mikrocomputer.Programmiersprache.Basic
42)	CASCADING_THEMES	Mikrocomputer.Programmiersprache.Basic. Hersteller.SystemSoft
46)	SPLITTING_THEMES	Mikrocomputer.Systemsoftware.Pascal-Compiler
	SPLITTING_THEMES	Mikrocomputer.Programmiersprache.Pascal
49)	CASCADING_THEMES	Mikrocomputer.Systemsoftware.Pascal-Compiler. Hersteller.PasWare
	CASCADING_THEMES	Mikrocomputer.Programmiersprache.Pascal. Hersteller.PasWare

The word expert just discussed accounts for a single frame (here: Mikrocomputer) with nested slot values of arbitrary depth. This basic description only slightly has to be changed to account for knowledge structures which are implicitly connected in the text. Basically divergent types of coherence patterns are worked out by word experts operating on, e.g. aspectual or contrastive coherence relations (cf. HAHN 1984).

#### 4. The Generation of Text Graphs Based on Topic/Comment Monitoring

The procedure of text graph generation for this basic type of thematic progression can be described as follows. After initialization by drawing upon the first frame entry occurring in co-text the text graph gets incrementally constructed whenever a new coherence reading is available in the corresponding data repository. Then, it has to be determined, whether its first parameter equals the current node of text graph which is either the leaf node of the initialized text graph (when the procedure starts) or the leaf node of the topic/comment subgraph which has previously been attached to the text graph. If equality holds, the coherence reading is attached to this node of the graph (including some merging operation to exclude redundant information from the text graph). If equality does not hold, remaining siblings or ancestors (in this order) are tried, until a node equal to the first parameter of the current coherence reading is found to which the reading will be attached directly. If no matching node in the text graph can be found, a new text graph is constructed which gets initialized by the current coherence reading. The text graph as the result of parsing of the text segment [1] with respect to the coherence readings generated in sec.3.2 is provided in App.C.

Note that the text graph generation procedure allows for an interpretation of basic coherence readings supplied by various word experts in terms of compound patterns of thematic progression, e.g. as given by the exposition of splitting rhemes (DANES 1974). Nevertheless, the whole procedure essentially depends upon the continuous availability of reference topics to construct a

coherent graph. Accordingly, the graph generation procedure also operates as a kind of topic/comment monitoring device. Obviously, one also has to take into account defective topic/comment patterns in the text under analysis. The SEPARATOR reading is a basic indicator of interruptions of topic/comment sequencing. Its evaluation leads to the notion of topic/comment islands for texts which only partially fulfill the requirements of topic/comment sequencing. Further coherence readings are generated by computations based solely on world knowledge indicators generating condensed lists of dominant concepts (lists of topics instead of topic graphs) (HAHN/REIMER 1984).

## 5. Conclusion

In this paper we have argued in favor of a word expert approach to text parsing based on the notions of text cohesion and text coherence. Readings word experts work out are represented in text graphs which illustrate the topic/comment structure of the underlying texts. Since these graphs represent the texts' thematic structure they lend themselves easily for abstracting purposes. Coherency factors of the text graphs generated, the depth of each text graph, the amount of actual branching as compared to possible branching, etc. provide overt assessment parameters which are intended to control abstracting procedures based on the topic/comment structure of texts. In addition, as much effort will be devoted to graphical modes of system interaction, graph structures are a quite natural and direct medium of access to TOPIC as a text information system.

## ACKNOWLEDGEMENTS

I would like to express my deep gratitude to U. Reimer for many valuable discussions we had on the word expert system of TOPIC. R. Hammwoehner and U. Thiel also made helpful remarks on an earlier version of this paper.

## REFERENCES

- Critz, J.T.: Frame Based Recognition of Theme Continuity. In: COLING 82: Proc. of the 9th Int. Conf. on Computational Linguistics. Prague: Academia, 1982, pp.71-75.
- Danes, F.: Functional Sentence Perspective and the Organization of the Text. In: F. Danes (ed): Papers on Functional Sentence Perspective. The Hague, Paris: Mouton, 1974, pp.106-128.
- Dijk, T.A. van: Text and Context: Explorations in the Semantics and Pragmatics of Discourse. London, New York: Longman, (1977) 1980 (a).
- Dijk, T.A. van: Macrostructures: An Interdisciplinary Study of Global Structures in Discourse, Interaction, and Cognition. Hillsdale/NJ: L. Erlbaum, 1980 (b).

Grimes, J.E.: Topic Levels. In: TINLAP-2: Theoretical Issues in Natural Language Processing-2. New York: ACM, 1978, pp.104-108.

Hahn, U.: Textual Expertise in Word Experts: An Approach to Text Parsing Based on Topic/Comment Monitoring (Extended Version). Konstanz: Univ. Konstanz, Informationswissenschaft, (May) 1984 (= Bericht TOPIC-9/84).

Hahn, U. & Reimer, U.: Word Expert Parsing: An Approach to Text Parsing with a Distributed Lexical Grammar. Konstanz: Univ. Konstanz, Informationswissenschaft, (Nov) 1983 (= Bericht TOPIC-6/83). [In: Linguistische Berichte, No.88, (Dec) 1983, pp.56-78. (in German)]

Hahn, U. & Reimer, U.: Computing Text Constituency: An Algorithmic Approach to the Generation of Text Graphs. Konstanz: Univ. Konstanz, Informationswissenschaft, (April) 1984 (= Bericht TOPIC-8/84)).

Halliday, M.A.K. / Hasan, R.: Cohesion in English. London: Longman, 1976.

Hobbs, J.R.: Coherence and Coreference. In: Cognitive Science 3. 1979, No.1, pp.67-90.

Hobbs, J.R.: Towards an Understanding of Coherence in Discourse. In: In: W.G. Lehnert / M.H. Ringle (eds): Strategies for Natural Language Processing. Hillsdale/NJ, London: L. Erlbaum, 1982, pp.223-243.

Reimer, U. & Hahn, U.: A Formal Approach to the Semantics of a Frame Data Model. In IJCAI-83: Proc. of the 8th Int. Joint Conf. on Artificial Intelligence. Los Altos/CA: W. Kaufmann, 1983, pp.337-339.

Small, S. / Rieger, C.: Parsing and Comprehending with Word Experts (a Theory and its Realization). In: W.G. Lehnert / M.H. Ringle (eds): Strategies for Natural Language Processing. Hillsdale/NJ: L. Erlbaum, 1982, pp.89-147.

