

## TEMPORAL INFERENCES IN MEDICAL TEXTS

Klaus K. Obermeier

Battelle's Columbus Laboratories  
505 King Avenue  
Columbus, Ohio 43201-2693, USA

### ABSTRACT

The objectives of this paper are twofold, whereby the computer program is meant to be a particular implementation of a general natural language [NL] processing system [NLPS] which could be used for different domains. The first objective is to provide a theory for processing temporal information contained in a well-structured, technical text. The second objective is to argue for a knowledge-based approach to NLP in which the parsing procedure is driven by extra linguistic knowledge.

The resulting computer program incorporates enough domain-specific and general knowledge so that the parsing procedure can be driven by the knowledge base of the program, while at the same time employing a descriptively adequate theory of syntactic processing, i.e., X-bar syntax. My parsing algorithm not only supports the prevalent theories of knowledge-based parsing put forth in AI, but also uses a sound linguistic theory for the necessary syntactic information processing.

### 1.0 INTRODUCTION

This paper describes the development of a NLPS for analyzing domain-specific as well as temporal information in a well-defined text type. The analysis, i.e. output, of the NLPS is a data structure which serves as the input to an expert system. The ultimate goal is to allow the user of the expert system to enter data into the system by means of NL text which follows the linguistic conventions of English.

The particular domain chosen to illustrate the underlying theory of such a system is that of medical descriptive texts which deal with patients' case histories of liver diseases. The texts are taken unedited from the *Journal of the American Medical Association*. The information contained in those texts serves as input to PATREC, an intelligent database assistant for MDX, the medical expert system [Chandrasekaran 83]. The objectives of this research are twofold, whereby the system described above is meant to be a particular implementation of a general NLP which could be used for a variety of domains.

The first objective is to provide a theory for processing temporal information contained

in a given text. The second objective is to argue for a knowledge-based approach to NL processing in which the parsing procedure is driven by extra linguistic knowledge.

My NLPS, called GROK, [Grammatical Representation of Objective Knowledge] is a functioning program which is implemented in ELISP and EFRL on a DEC20/60. The full documentation, including source code is available [Obermeier 84]. The program performs the following tasks: (1) parse a text from a medical journal while using linguistic and extra linguistic knowledge; (2) map the parsed linguistic structure into an event-representation; (3) draw temporal and factual inferences within the domain of liver diseases; (4) create and update a database containing the pertinent information about a patient.

### 2.0 OVERVIEW

#### 2.1 A Sample Text

The user of my NLPS can enter a text of the format given in Figure 1<sup>1</sup>. The texts which the NLPS accepts are descriptive for a particular domain. The information-processing task consists of the analysis of linguistic information into datastructures which are chronologically ordered by the NLPS.

1. This 80-year-old Caucasian female complained of nausea, vomiting, abdominal swelling, and jaundice.
2. She had diabetes mellitus, treated with insulin for six years before admission.
3. She had had ill-defined gastrointestinal complaints for many years and occasional episodes of nausea and vomiting three years previously.
4. Four weeks before admission she developed pain across the upper abdomen, radiating to the flanks.
5. She also complained of shooting precordial pains and palpitation with slight exertional dyspnea.

Figure 1: Sample Text for Case No. 172556

---

<sup>1</sup>The numbering on the sentences is only for ease of references in the following discussion and does not appear in the actual text.

The first module of the program analyzes each word by accessing a lexical component which assigns syntactic, semantic, and conceptual features to it. The second module consists of a bottom-up parser which matches the output from the lexical component to a set of augmented phrase structure rules<sup>2</sup>. The third module consists of a knowledge base which contains the domain-specific information as well as temporal knowledge. The knowledge base is accessed during the processing of the text in conjunction with the augmented phrase structure rules.

The output of the program includes a lexical feature assignment as given in Figure 2, a phrase-structure representation as given in Figure 3, and a knowledge representation as provided in Figure 4. The resulting knowledge representation of mv NLPS consists of a series of events which are extracted from the text and chronologically ordered by the NLPS based on the stored knowledge the system has about the domain and general temporal relations. The final knowledge representation (see Figure 5) which mv NLPS generates is the input to the expert system or its database specialist. The final output of the expert system is a diagnosis of the patient.

```
THIS DET: EIGHTY-YEAR-OLD ADJ AGE;
          CAUCASIAN ADJ RACE;
          FEMALE N SEX;
          COMPLAIN V
          ((GET-NEED-NEW (QUOTE COMPLAIN)
                        (QUOTE CLASSIF)
                        (QUOTE $VALUE)))
          (ED))
          (OF PREP)
          (NAUSEA N SIGNSYMPOM)
          *COMMA* BOUNDARY
          VOMIT N V SIGNSYMPOM (ING))
          *COMMA* BOUNDARY
          ABDOMINAL ADJ
          SWELL V SIGNSYMPOM (ING))
          *COMMA* BOUNDARY
          AND BOUNDARY
          JAUNDICE N SIGNSYMPOM
          *PERIOD* BOUNDARY
```

Figure 2: Lexical Access of Sentence (1) in Figure 1

## 2.2 Scenario

The comprehension of a descriptive text requires various types of knowledge: linguistic knowledge for analyzing the structure of words and sentences; "world knowledge" for relating the text to our experience; and, in the case of technical texts, expert knowledge for dealing with information geared toward the domain expert. For the purpose of mv research, I contend that the comprehension of technical, descriptive text is simply a conversion of information from one representation into another based on the knowledge of the NLPS.

<sup>2</sup>The augmentation consists of rules which contain knowledge about morphology, syntax, and the particular domain in which the NLPS is operating. These rules are used for interpreting the text, in particular, ambiguities, as well as for generating the final output of the NLPS.

```
[[N2]
  ((DET) THIS DET) ((ADJ) EIGHTY-YEAR-OLD ADJ AGE)
  ((ADJ) ((ADJ) CAUCASIAN ADJ RACE))
  ((N) ((N) FEMALE N SEX))
  ..NP: head: FEMALE

((V)
  ((V-PART) ((V) COMPLAIN
             V
             ((GET-NEED-NEW (QUOTE COMPLAIN)
                           (QUOTE CLASSIF)
                           (QUOTE $VALUE)))
             (ED))
  ((PART) OF PART))
  ..VP: the ed-suffix is separated;
  the trigger on complain changed
  the following of from a preposition
  to a particle:

  ((N) NAUSEA N SIGNSYMPOM))
  ..this N is part of the VP

*COMMA* BOUNDARY)
  ..punctuation breaks up phrases

[[N2] ((N) ((N) VOMIT N SIGNSYMPOM (ING)))
  ..the noun/verb
  ambiguity on this word has been
  resolved by the "ING-Specialist"

*COMMA* BOUNDARY)

[[N2] ((ADJ) ((ADJ) ABDOMINAL ADJ)) ((N) ((N) SWELL (ING) N))
  .."ING" changed the verb to a gerund

*COMMA* BOUNDARY)

*AND BOUNDARY)

[[N2] ((N) ((N) JAUNDICE N SIGNSYMPOM))

*PERIOD* BOUNDARY)
```

Figure 3: Syntactic Annotation for Sentence (1) in Figure 1

```
EVENT1
SYMPTOM: NAUSEA/VOMIT/ABDOMINAL/SWELLING/JAUNDICE
KEY EVENT: ADMISSION
DURATION: ADMISSION

EVENT2
SYMPTOM: DIABETES MELLITUS
KEY EVENT: ADMISSION
RELATION TO KEY EVENT: 6 YEARS BEFORE
DURATION: SIX YEARS

EVENT3
SYMPTOM: GASTROINTESTINAL COMPLAINT
KEY EVENT: ADMISSION
RELATION TO KEY EVENT: MANY YEARS
DURATION: MANY YEARS

EVENT4
SYMPTOM: NAUSEA/VOMIT
KEY EVENT: ADMISSION
RELATION TO KEY EVENT: 3 YEARS BEFORE
DURATION: INTERMITTENT
```

Figure 4: A Simplified Sample Output of the Representation of Sentences (1), (2), and (3) from Figure 1

<sup>3</sup>This partial parse of the sentence follows Jackendoff's X-bar theory [Jackendoff 77], which is discussed in [Obermeier 84, 85]; roman numerals indicate the number of bars assigned to each phrase. Comments to the parse were made after the actual run of the program.

If a doctor were given a patient's case history (see Figure 1), he would read the text and try to extract the salient pieces of information which are necessary for his diagnosis. In this particular text type, he would be interested in the sign, symptoms, and laboratory data, as well as the medical history of the patient. The crucial point hereby is the temporal information associated with the occurrences of these data. In general, he would try to cluster certain abnormal manifestations to form hypotheses which would result in a coherent diagnosis. The clustering would be based on the temporal succession of the information in the text. Each manifestation of abnormalities I will refer to as an "event". Each event is defined and related to other events by means of temporal information explicitly or implicitly provided in the text. An important notion which I use in my program is that of a key event<sup>4</sup>. "Events are organized around key events (which are domain-specific - in the medical domain, some of the important ones are 'admission', 'surgery', 'accident', etc.), so that other events are typically stated or ordered with respect to these key events" [Mittal 82].

### 3.0 KNOWLEDGE-BASED PARSING

#### 3.1 Selection and Organization for the Knowledge Base

I have characterized the task of a doctor reading a patient's case history as finding key domain concepts (e.g., sign, symptom, laboratory data), relating them to temporal indicators (e.g., seven years ago), and ordering the events resulting from assigning temporal indicators to key concepts with respect to a "key event" (e.g., at admission, at surgery).

(1) This 80-year-old Caucasian female complained of nausea, vomiting, abdominal swelling and jaundice.

In the sample text in Figure 1, the first sentence, given in (1) requires the following domain concepts:

Patient: person identified by age, sex, and profession, whose signs, symptoms, and laboratory data will be given.

Symptoms: manifestations of abnormalities reported by the patient. Certain symptoms have to be further defined: swelling needs a characterization as to where it occurs. Pain can be characterized by its location, intensity, and nature (e.g., "shooting").

Signs: abnormalities found by the physician such as fever, jaundice, or swelling.

<sup>4</sup>The notion of "key event" is further discussed in 4.3 "Key Events".

Whether "fever" is a sign or a symptom is indicated by the verb. Therefore, the verbs have features which indicate if the following is a sign or a symptom. There are no explicit temporal indicators in (1), except the tense marker on the verb. The doctor, however, knows that case histories ordinarily use "admission" as a reference point.

```

/PASSERT EVNT:
  /SYMPOM /SVALUE (|||||NIL NAUSEA|VOMIT|ABDOMINAL SWELL|JAUNDICE|
  /KEY-EVENT /SVALUE /ADMISSION)))
  /DURATION /SVALUE /ADMISSION)))
  /CLASSIF /SVALUE /INDIVIDUAL)))
  /TYPE /SVALUE /EVENT)))

```

Figure 5: Final Knowledge Representation of Event 1 in EFRL

(2) She had diabetes mellitus, treated with insulin for six years before admission.

The sentence in (2) requires a temporal concept "year" in conjunction with the numerical value "six", it also requires the concept "duration" to represent the meaning of for. The "key event" at admission is mentioned explicitly and must be recognized as a concept by the system.

After selecting the facts on the basis of about 35 case descriptions as well as previous research of the medical sublanguage [Hirschman 83]<sup>5</sup>, I organized them into schemas based on what is known about the particular text type. In [Bonnet 79], a medical summary is characterized as "a sequence of episodes that correspond to phrases, sentences, or groups of sentences dealing with a single topic. These constitute the model and are represented by schemas" [Bonnet 79, 80]. Schemas for the medical domain in Bonnet's system are SPATIENT-INFORMATION (e.g., sex, job), SSIGNS (e.g., fever, jaundice). In GROK, I use the schemas SREPORT-SIGN, SREPORT-SYMPOM, SREPORT-LAB-DATA, SPATIENT-INFO. Each of my schemas indicates "who reports, what to whom, and when". The SREPORT-SYMPOM schema has the following elements: verb(unknown), subject(patient), object-(symptom), indirect object(medic), time(default is admission).

After selecting the facts on the basis of the domain, and organizing them on the basis of the text-type, I add one fact for putting the information into the target representation. The target representation consists of a temporal indicator attached to a domain-specific fact - what I had referred to in as "event". The event structure contains the following elements: name of domain-specific concept, reference point, duration (known or unknown), and relation to reference point (e.g., before, after).

<sup>5</sup>I use ten types of domain-specific facts: sign, symptom, lab data, body-part, etc., I use six temporal facts: month, year, day, week, duration, period, i.e., "for how long".

### 3.2 The Flow of Control

In addition to domain-specific knowledge, a person reading a text also uses his linguistic knowledge of the English grammar. The problem for a NLPS is how to integrate linguistic and extra linguistic knowledge. The dominant paradigm in computational linguistics uses syntactic and morphological information before considering extra linguistic knowledge; if extra linguistic knowledge is used at all.

Considering syntactic knowledge before any other type of knowledge has the following problems which are avoided if enough contextual information can be detected by the knowledge base of the NLPS:

- global ambiguities cannot be resolved (e.g., Visiting relatives can be boring)
- word-class ambiguities (e.g., bank) and structural ambiguities cause multiple parses (e.g., I saw the man on the hill with the telescope).

Moreover, psycholinguistic experiments have shown [Marslen-Wilson 75, Marslen-Wilson 78, Marslen-Wilson 80] that the syntactic analysis of a sentence does not precede higher level processing but interacts with semantic and pragmatic information. These findings are, to some extent, controversial, and not accepted by all psycholinguists.

In my system, knowledge about the domain, the text-type, and the target representation is used before and together with syntactic information. The syntactic information helps to select the interpretation of the sentence. Syntax functions as a filter for processing information. It selects the constituents of a sentence, and groups them into larger "chunks", called phrases. The phrase types noun phrases [NP] and verb phrase [VP] contain procedures to form concepts (e.g., "abdominal pain"). These concepts are combined by function specialists. Function specialists consists of procedures attached to function words (e.g., prepositions, determiners), inflectional morphemes, and boundary markers (e.g., comma, period).

Technically, I distinguish between phrase specialists and function specialists. The phrase specialists interact with extralinguistic knowledge to determine which concepts are expressed in a text, the function specialists determine locally what relation these concepts have to each other. So in general, the phrase specialists are activated before the function specialists.

To illustrate this process, consider the sentence:

(3) The patient complained of shooting pain across the flanks for three days before admission.

The NP-specialist combines the and patient into a phrase. The central processing component in the sentence is the VP-specialist. Its task is to find the verb-particle construction (complain of), and the object (e.g., shooting pain). The VP-specialist also looks at the syntactic and semantic characteristics of complain of. It notes that complain of expects a symptom in its object position. The expectation of a symptom invokes the schema "report-symptom". At this point, the schema could fill in missing information, e.g., if no subject had been mentioned, it could indicate that the patient is the subject. The schema identifies the current topic of the sentence, viz., "symptom".

GROK next encounters the word shooting. This word has no further specification besides that of being used as an adjective. The head noun pain points to a more complex entity "pain" which expects further specifications (e.g., location, type). It first tries to find any further specifications within the analyzed part of the NP. It finds shooting and adds this characteristic to the entity "pain". Since "pain" is usually specified in terms of its location, a place adverbial is expected. Upon the entry of across, the entity "pain" includes "across" as a location marker, expecting as the next word a body-part. The next word, flank is a body-part, and the "pain" entity is completed. Note here, that the attachment of the preposition was guided by the information contained in the knowledge base.

The next word for is a function word which can indicate duration. To determine which adverbial for introduces, the system has to wait for the information from the following NP-specialist. After the numeric value "three", the temporal indicator "day" identifies for as a duration marker.

Explicit temporal indicators such as day, week, or month, under certain conditions introduce new events. As soon as GROK verifies that a temporal indicator started an event, it fills in the information from the "report-xxx" schema. The new event representation includes the sign, symptom, or laboratory data, and the temporal indicator. The last two words in the sample sentence before admission, provide the missing information as to what "key event" the newly created event is related to.

Once a new event frame or domain-specific frame is instantiated, GROK can use the information associated with each event frame (e.g., duration, key-event), together with the information from the domain-specific frame (e.g., the pain frame contains slots for specifying the location, intensity, and type of pain) to interpret the text.

## 4.0 TEMPORAL INFORMATION PROCESSING

### 4.1 Problems

The inherent problems of text comprehension from an information processing viewpoint are how to deal with the foremost problems in computational NLP (e.g., ambiguity, anaphora, ellipsis, conjunction), including the foremost problems in temporal information processing (e.g., implicit time reference, imprecision of reference).

Within AI and computational linguistics, only a few theories have been proposed for the processing of temporal information [Kahn 77, Hirschman 81, Kamp 79, Allen 83]. In particular, a theory of how a NLP can comprehend temporal relations in a written text is still missing. In my research, I present a theory for processing temporal information in a NLPS for a well-defined class of technical descriptive texts. The texts deal with a specific domain and tasks which require the processing of linguistic information into a chronological order of events. The problems for processing the temporal information contained in the text include:

- a NLPS has to work with implicit temporal information. Although in (1), no explicit temporal reference is present, the NLPS has to detect the implied information from the context and the extra linguistic knowledge available.
- a NLPS has to work with fuzzy information. The reference to for many years in (3) is fuzzy, and yet a NLPS has to relate it to the chronology of the case.
- a NLPS has to order the events in their chronology although they are not temporally ordered in the text.

### 4.2 Solutions

My solution to the problems discussed in the previous section lies within the computational paradigm as opposed to the Chomskyan generative paradigm. The computational paradigm focuses on how the comprehension processes are organized whereas within the generative paradigm, linguistic performance is of less importance for a linguistic theory than linguistic competence. Within the computational paradigm, the representation and use of extra-linguistic knowledge is a major part of studying linguistic phenomena, whereas generative linguists separate linguistic phenomena which fall within the realm of syntax from other cognitive aspects [Winograd 83, 21].

Functionality is the central theoretical concept upon which the design of GROK rests. What is important for comprehending language is the function of an utterance in a given situation. Words are used for their meaning, and the meaning depends on the use in a given context. The meaning of a word is subject to change according to the context, which is based on the function of the words that make up the text. Therefore, my approach to building a NLPS focuses on modeling the context of a text in a particular domain. I am primarily concerned with the relationship between writer-text-reader, rather than with the relationship between two sentences. The use of the context for parsing requires a knowledge representation of the domain, and the type of text, in addition to linguistic and empirical knowledge.

In contradistinction to NLPSs which use syntactic information first [Thompson 81], and which possibly generate unnecessary structural descriptions, my system uses higher level information (e.g., domain, text-type) before and together with usually a smaller amount of syntactic information. In GROK, the syntactic information selects between contextually interpretations of the text - *syntax acts as a filter* for the NLPS.

In contradistinction to NLPSs which use conceptual information first [Schank 75], GROK, partially due to the limited information processing task and the particular domain, starts out with a small knowledge base and builds up datastructures which are used subsequently in the processing of the text. The knowledge base of my system contains only the information it absolutely needs, whereas Schankian scripts have problems with when to activate scripts and when to exit them.

### 4.3 Key Events

Temporal information in a text is conveyed by explicit temporal indicators, implicit temporal relations based on what one knows about written texts (e.g., "time moves forward"), and "key events". I define a key event as a domain-specific concept which is used to order and group events around a particular key event. In my theory, temporal processing is based on the identification of key events for a particular domain, and their subsequent recognition by the NLPS in the text.

Temporal indicators in a sentence are not of equal importance. The tense marking on the verb has been the least influential for filling in the event structure. For the program, the most important sources are adverbials.

The linear sequence of sentences also contributes to the set-up of the configurations of events. My program makes use of two generally known heuristics; time moves forward in a narrative if not explicitly stated otherwise;

the temporal reference of the subordinate clause is ordinarily the same as that in the main clause.

"Key events" are significant since they are used to relate events to one another. In my theory of text processing, key events build up the temporal structure of a text. If key events for other domains can be identified, they could be used to explain how a NLPs can "comprehend" the texts of the domain in question.

The representation of temporal information is significant in my theory. I define an event as the result of the assignment of a temporal value to a domain-specific concept. The structure of an event is generalizable to other domains. An event consists of a domain-specific concept, a key event, a relation to key event, and a duration. In the medical domain, the instantiated event contains information about how long, and when a symptom or sign occurred, and what the key event of the instantiated event was.

Apart from the temporal issue, my research has shown that if the domain and the task of the NLPs are sufficiently constrained, the use of frames as a knowledge representation scheme is efficient in implementing GROK. In my program, I have used individual frames to represent single concepts (e.g., pain). These concepts help the NLPs to access the domain-specific knowledge base. Together with the temporal indicators, the information from the knowledge base is then transferred to the topmost event frame. Procedures are then used to relate various event frames to each other. The restrictions and checks on the instantiation of the individual frames preclude an erroneous activation of a frame.

The viability of this approach shows that the idea of stereotypical representation of information is useful for NLPs if properly constrained. My program checks for the accessibility of the various levels of the knowledge representation whenever new information is coming in. This multilayer approach constrains the instantiation of the event frame sufficiently in order to prevent erroneous event instantiation.

#### 4.4 Comparison to Extant Theories on Temporal Processing

The overall ideas of GROK as they relate or differ from the extant theories and systems are introduced by looking at four major issues concerning temporal processing.

- temporality: how is an event defined in the system; how is temporal information treated vis-a-vis the whole system? What search algorithms or inference procedures are provided?

- organization: are events organized on a time line, by key events, calendar dates, before/after chains?
- problems: how is imprecision, fuzziness, and incompleteness of data handled?
- testing: how can the system be tested; by queries, proofs, etc.? Does it have a consistency checker?

In GROK, I use an interval-based approach to temporal information processing. An event is defined as an entity of finite duration. As in [Kamp 79, 377], event structures are transformed into instants by the Russell-Wiener construction.

In GROK, the NLPs processes temporal information by first associating a concept with a temporal reference, then evaluating the extension of this event. The evaluation considers syntactic (e.g., adverbials) and pragmatic information (current time focus). Each event is represented in the knowledge base with information about when, for how long, and what occurred.

The parser while analyzing the sentences, orders these events according to a "key event". The single events contain information about the temporal indicator which is attached to a domain-specific fact. The single events are connected to the respective "key event". "Key events" are domain-specific. In general, I stipulate that every domain has a limited number of such "key events" which provide the "hooks" for the temporal structure of a domain-specific text.

GROK also differs from logical theories in that it deals with discourse structures and their conceptual representations, not with isolated sentences and their truth value. It is different from Kahn's time specialist [Kahn 77] in that it uses domain knowledge and "knows" about temporal relations of a particular domain. Moreover, Kahn's program only accepts LISP-like input and handled only explicit temporal information. The use of domain-specific temporal knowledge also sets GROK apart from Allen's [Allen 83] temporal inference engine approach.

GROK differs from Kamp's discourse structures in that it uses the notion of reference intervals that are based on conventional temporal units (e.g., day, week, month, year) to organize single events into chronological order.

GROK is in many respects similar to research reported in [Hirschman 1981]: both systems deal with temporal relations in the medical domain; both systems deal with implicit and explicit temporal information. GROK differs

from Hirschman's system in that GROK uses domain-specific and other extra linguistic information for analyzing the text, whereas Hirschman relies primarily on available syntactic information. Therefore, Hirschman's system as presented in [Hirschman 81] can neither handle anaphoric references to continuous states nor represent imprecision in time specification.

#### 4.5 State of Implementation

GROK is a highly exploratory program. The limitations of the current implementation are in three areas:

- The parser itself does not provide the capability of a chart parser since it will not give different interpretations of a structurally ambiguous sentences. This type of structural ambiguity, where one constituent can belong to two or more different constructions, would not be detected.
- The knowledge base does not have a fully implemented frame structure. Each generic frame has a certain number of slots that define the concept. A generic concept (e.g., sign) must have slots which contain possible attributes of the specific frame (e.g., where is the sign found; how severe is its manifestation). These slots have not yet been implemented. The number of frames is strictly limited to the temporal frames and a few exemplary generic frames necessary to process the text.
- The range of phenomena is limited. Only "before-admission" references are recognized by the system. Furthermore, slots that prevent the inheritance of events of limited durations are not yet in place.

In general, GROK is still in a developmental stage at which a number of phenomena have yet to be accounted for through an implementation.

#### 5.0 CONCLUSION

In this paper, I argued for an integration of insights gained from linguistic, psychological, and AI-based research to provide a pragmatic theory and cognitive model of how temporal inferences can be explained within the framework of computational information processing. A pragmatic theory focuses on the information from the context (e.g., co-text, discourse situation, intentions of interlocutors) to explain linguistic behavior.

I have shown how an integration of linguistic and extra linguistic knowledge achieves a form of comprehension, where comprehension is characterized as a conversion of information based on knowledge from one representation into another. I have also shown how this approach leads to a parsing technique which avoids common pitfalls, and, at the same time, is consistent with results in psycholinguistic research. I have furthermore shown that such a procedural approach is a basis for an event-based theory for temporal information processing.

In particular, the findings implemented in GROK show the shortcomings of the orthodox rule-based approach to language processing which reduces words to tokens in a larger context while overemphasizing the role of the phrase and sentence level. It does this by providing a temporal knowledge representation and algorithms for processing pragmatic information which are applicable to a wider range of phenomena than most of the notable computational NL theories within the field of AI [Schank 81, Rieger 79, Wilks 75], or linguistics [Marcus 80].

In particular, my research shows that:

- NL can be processed realistically by a deterministic algorithm which can be interpreted in a mental model. A realistic NLP system tries to emulate human behavior. A deterministic parser works under the assumption that (1) a human NLP system makes irrevocable decisions during processing and (2) that humans are not unconstrained "wait-and-see-parsers" [Kac 82]. A mental model provides an internal representation of the state of affairs that are described in a given sentence [Johnson-Laird 81].
- Temporal information processing is adequately explained only in a pragmatic theory that captures the duality of interval and point-based representation of time. In my theory, temporal processing is possible because of domain-specific key events which provide the "hooks" for the temporal structure of a text.
- NL can be processed efficiently by a set of integrated linguistic and extra linguistic knowledge sources.

## REFERENCES

- [Allen 83]  
Allen, J.F.  
Maintaining Knowledge about Temporal Intervals  
CACM 26, 1983.
- [Bonnet 79]  
Bonnet, A.  
Understanding Medical Jargon as if it were Natural Language.  
In Proc from IJCAI 6. 1979.
- [Chandrasekaran 83a]  
Chandrasekaran, B. and Mittal, S.  
Conceptual Representation of Medical Knowledge for Diagnosis by Computer: MDX and Associated Systems  
1983  
Advances in Computer, Vol. 22.
- [Hirschman 81]  
Hirschman, L., Story, G.  
Representing implicit and explicit time relations in narrative.  
in IJCAI 81. 1981
- [Hirschman 83]  
Hirschman, L. and Sager, N.  
Automatic Information Formatting of a Medical Sublanguage.  
In Kittredge (editor), Sublanguage. deGruyter, 1983
- [Johnson-Laird 81]  
Johnson-Laird, P.N.  
Mental Model of Meaning.  
In Joshi, A. Webber, B. Sag, I (editor), Elements of Discourse Understanding. Cambridge University Press, 1981.
- [Kac 82]  
Kac, M.B.  
Marcus: A theory of syntactic recognition for NL (Review).  
Language 58:447-454, 1982.
- [Kahn 77]  
Kahn, K. and Corry, G.A.  
Mechanizing Temporal Knowledge.  
Artificial Intelligence 9, 1977.
- [Kamp 79]  
Kamp, H.  
Events, Instants and Temporal Reference.  
In Baeurle, R., Egli, U., Stechow, A. (editors), Semantics from Different Points of View, Springer, 1979.
- [Marcus 80]  
Marcus, M.  
Theory of Syntactic Recognition for Natural Language.  
MIT Press, 1980.
- [Marslen-Wilson 75]  
Marslen-Wilson, W.D.  
Sentence perception as in interactive parallel process.  
Science 189, 1975.
- [Marslen-Wilson 78]  
Marslen-Wilson, W. and Welsh, A.  
Processing interactions and lexical access during word recognition in continuous speech.  
Cognitive Psychology 10, 1978.
- [Marslen-Wilson 80]  
Marslen-Wilson, W. and Tyler, L.  
The temporal structure of spoken language understanding; the perception of sentences and words in sentences.  
Cognition 8, 1980.
- [Mittal 82]  
Mittal, S.  
Event-based Organization of Temporal Databases.  
1982  
Processings from the 4th National Conference of the Canadian Society for Computational Studies of Intelligence, Saskaton, Canada.
- [Obermeier 84]  
Obermeier, K.  
Temporal Inferences in Computational Linguistic Information Processing.  
The Ohio State University, Ph.D. Dissertation, 1984.
- [Obermeier 85]  
Obermeier, K.  
Grok - a natural language front end for medical expert systems.  
In Proceedings from the 5th International Workshop on Expert Systems and Their Applications.  
Palais des Papes - Avignon, France  
May 13-15, 1985.
- [Rieger 79]  
Rieger, C. and Small, S.  
Word Expert Parsing  
6th IJAI, 1979.
- [Schank 75]  
Schank, R.  
Conceptual Information Processing  
North Holland, 1975.
- [Schank 81]  
Schank, R.C. and Riesbeck, C.K.  
Inside Computer Understanding. Five Programs Plus Miniatures.  
Lawrence Erlbaum Associates, 1981.
- [Thompson 81]  
Thompson, H.  
Chart Parsing and rule schemata in PSG.  
1981  
19th Ann. ACL.



[Wilks 75]

Wilks, Y.  
An intelligent analyzer and understander  
of English.  
CACM 18, 1975.

[Winograd 83]

Winograd, T.  
Language as a Cognitive Process.  
Addison-Wesley, 1983.