Study of Time Structure Pattern in News Stories

Yunlong Jiang

A Major Report

in

The Department

of Computer Science

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Computer Science at
Concordia University
Montreal, Quebec, Canada

April 2004

Your file Votre référence
ISBN: 0-612-94764-5
Our file Notre référence
ISBN: 0-612-94764-5

ABSTRACT

Study of Time Structure Pattern in News Stories

Yunlong Jiang

News articles are a particular type of story that reflects our daily life on various media. The most distinctive characteristics of news story are its discontinuity in the ordering of events and its limited shelf life. These characteristics present difficulties for automated systems which deal with temporal information. Such applications include information extraction, question answering, summarization, machine translation, etc. In recent years, some concrete work has been done in the area of event identification and temporal expression annotation. Some issues still remain regarding implicit temporal information.

This major report describes a method to find patterns in news stories that journalists use to organize events. The study will help us better understand the nature of news stories. It may also help to get implicit temporal information by the implication from the event in the same paragraph. The major report also presents an annotation scheme for annotating temporal expressions and paragraphs in new texts. In this study, we manually annotated all temporal expressions and paragraphs in selected news texts. We build the time structure for each of the texts according to the story time of the paragraphs. After all the target news texts are analyzed and annotated, we compare time structures against each other to derive time structure patterns that may exist in news stories.

ACKNOWLEDGEMENTS

It has been both my pleasure and my good fortune to perform this work under the direction of Dr. Sabine Bergler. Her enthusiastic supervision has provided many stimulating discussions and thoughtful suggestions. At the same time she always permitted me the freedom to work and think independently. I would like to take this opportunity to thank Dr. Bergler for her encouragement, patience, guidance and kindness.

I am grateful to my colleague Zhuoyan Li for his encouragement and inspiration. He has given much support to me throughout this major report.

Finally, my deepest appreciation goes to my wife for the love and understanding she has provided during the past few years. I am especially grateful to my children: Andrew, Michael and Samuel for making life enjoyable.

Table of Contents

LIST OF TABLES

Chapter 2: Literature Review

Chapter 3: Temporal Structure of News Stories

LIST OF FIGURES

Chapter 2: Literature Review

Chapter 3: Temporal Structure of News Stories

CHAPTER 1

INTRODUCTION

1.1 The Problem

The news is one special kind of stories that are told on various media such as television, broadcast and newspaper. These stories recap daily happenings of our societies and serve news consumers a variety of purposes, information, entertainment and persuasion, for example.

News stories differ from other kind of stories in terms of the way they are told. For almost all other kinds of stories, events tend to be told in chronological order, the order in which the events happened. This is proven the most natural way to tell a story because it matches its discourse structure to the event structure. It is independent from genre, language and culture. There is research on news discourse structure showing that stories are most comprehensible when events are told in the order they happened. Ohtsuka and Brewer (1992) found there is a significant drop in comprehension level for the same story that is described in reverse order. More drop occurs when the same story is presented using flashback for some of its events. Labov and Waletzky (1967) reported a study on personal narrative. The study shows that the event structures match their discourse structures in most narratives. If one changes the order of narrative clauses, one changes the order of events.

Although it is more logical to present events in chronological order than any other order, the news, however, go to the opposite direction. Linguists have noticed that events in news stories are often described in a sequence of non-chronological order. There exist discontinuities in the discourse structures of news stories. The central event of the news usually is introduced at the beginning, and then the news story returns to it for several times afterward. Discontinuity and non-chronology are the most characterized differences between news stories and narratives.

"Order is everything, but chronology is nothing" (Bell 1995a). Because the way news stories are presented, the times that associated with the events become crucial. If no times bind to the events in news stories, news stories would be no difference from narratives. The stereotype of the narratives would make news consumers think that all the events being told are in the order they happened. Apparently that is not the case. The discontinuity of the news stories makes news a little hard to comprehend.

For applications that temporal information is concerned, the discontinuity of discourse structure presents problems. Such applications include information extraction (e.g. normalizing temporal references for database entry), question answering (answering "when" questions), summarization (temporally ordering information), machine translation (translating and normalizing temporal references), and information visualization (viewing event chronologies). These applications have a common task, to extract temporal information, especially times that are related to events.

Another huge difference between news story and narrative is their life circle. Some narratives are valid or read for generations. But news is considered perishable. For news, "to know that something happened may be worthless without knowing when it happened" (Setzer 1991). For example, if one reads a news story of three months ago about how the United States raised its terror alert level from yellow to orange, one might not pay extra attentions to it because the alert level are already dropped back to yellow or even lower at the time. In fact even if the alert level remains high, the impact of the news is probably already weakened as time passes. Recency (Bell 1995a) is such an important factor in news value that it urges news agencies to pursue the goal of reporting events as soon as possible. Technological development now allows news stories to be told almost in real time. Breaking news are the ones that have most recency. They are usually the events that happened around world just moments ago.

Because of the news value of recency, news consumers usually have the presumption of which the news stories are the events just happened. This presumption changes the way that journalists construct news stories. Full, explicit time may not be given in headline, or in lead sentence, or even in text. Lack of full time expressions in news stories also presents difficulties in information extraction from news stories.

Answering questions about information in a news story after reading it is easy for humans because humans have the general knowledge, or common sense, about the world. Humans know how things or events are organized in real world no matter how they are ordered in a story, for example, a rescue work always comes after an accident happened and not other way around. For an automated system, answering the question "when did the rescue work start?" is not as easy as it is for humans. This involves identification of events and times in text, which is a very difficult job, and a very detailed analysis in events, times and their relationships.

In recent years, some concrete work has been done in the area of events identification and temporal expression annotation. The TIDES research group has developed a very thorough set of guidelines for annotating time expressions in text - TIDES Temporal Guidelines (Ferro et al. 2001). Setzer has developed a conceptual framework and annotation scheme (STAG – Sheffield Temporal Annotation Guidelines) for identifying and annotating events in text. The TERQAS workshop integrates both TIDES and STAG. The results from the workshop - TimeML Annotation Guidelines – provide a more thorough mean to annotate events, times, relations between events and relations between event and time. The applications in IE (information extraction) area are greatly benefiting from those research fruits. The frameworks and annotation schemes are discussed in more detail in chapter 2.

Despite significant progress in information extraction from text, some issues still remain. The frameworks mentioned above did not provide mechanisms for extracting implicit temporal information and implicit temporal relations. Thus extraction of such information still is a difficult part for those automated systems. On the other hand, the recall rate and precision of events and times recognition when applied to those annotation schemes are far from perfect compared to human annotator. Setzer (2001) did a pilot study to evaluate the scheme on a small corpus. A group of annotators were asked to use the scheme to identify events and times. She reported 77% recall rate and 81% precision on the events and times reported in the corpus compared to a human produced 'gold standard'. She also reported that the annotation scheme was difficult to apply without further improvement.

Besides those corpus-driven approaches, formal discourse approaches also provide frameworks and mechanisms for extracting temporal information from text. Labov and Waletzky (1967) developed a framework for the analysis of narratives of personal experience told in conversation. The framework focuses mainly on six elements: abstract, orientation, action, evaluation, resolution and coda.

Bell introduced a framework for analyzing the discourse structure of news (Bell 1995a; Bell & Garrett 1997). The framework provides guidelines for building structures for four aspects: events, time, place and news actors. The time structures built by using the framework are based on sentences. Every sentence is assigned an integer value indicating the order in which the events in the sentence happened.

1.2 The Objectives

A paragraph is a subdivision of a text intended to separate ideas. Each paragraph is a block of text that describes events that are closely related to each other. Sometimes events in the same paragraph happen in sequence. This characteristic suggests that knowing the time of one event in the paragraph may implicate the time of other events in the paragraph.

The goal of this project is to find any possible patterns that journalists use to organize events in their stories. This will help us better to understand the nature of news stories. Further we might get implicit temporal information that can't be caught by current IE application by the implication from the event in the same paragraph.

The study involves analyzing certain amount of news texts. For each news text, all temporal expressions and paragraphs are annotated manually. At the same time, a time structure is generated for further analysis. The method used to analyze time in news stories is similar to the one in Bell's framework (Bell 1995a). But in our study, the time structure of a news story is not based on sentences, rather it is built upon values of the paragraphs. After all the target news texts are analyzed and annotated, we compare time structures against each other to find if any pattern of time structure exists at certain percentage of all. The results are reported and analyzed in chapter 4.

We develop an annotation scheme for annotating paragraphs in news text. It simply assigns an identifier and a value to each paragraph. The annotation scheme is applied to the text after all the temporal expressions have been annotated. The value of a paragraph is dependent on the values of temporal expressions in the paragraph or determined according to the context.

The scheme for annotating temporal expressions is based on TimeML Annotation Guidelines. More detailed information about TimeML is described in section 2.4.3. Some modifications on tags and attributes are made for the needs of simplicity. Each temporal expression has a value in most explicit format of "yyyy-mm-dd-hh-nn-ss" that enables a straightforward comparison on it.

The target genre chosen is news stories because of its inclusion of temporal information and its accessibility for research. This genre has long been a focus amongst those working with language and communication including much of the information extraction work recently.

CHARPTER 2

LITERATURE REVIEW

2.1 Time

Time can be referred to as points or as intervals (Allen 1983). Consider the following expressions:

> *The blast occurred at 8:26 a.m.*

> *An explosion rocked the Royal Marines School of Music in a southeastern coastal town today.*

In the first sentence, "at 8:26 a.m." refers to precise points in time at which the blast occurred. The time expression "today" in second sentence refers to a time interval in which the explosion occurred.

These expressions have provided some sort of temporal precision that enables readers to determine "real" time on a calendar or clock. However, there are other temporal expressions describing temporal relations between the time intervals and can't be identified as corresponding to "real" time, e.g.:

> *The plane crashed after the pilot and his crew ejected.*

The temporal connective "after" only indicates the relationship between the time when plane crashed and the time when the crew ejected. No calendar time is associated to the expression.

A time point can be viewed as an instant in time line. In these approaches that use instants as the primitive units of times, time points are treated as atoms, the building blocks of time that are durationless and indivisible (Setzer 1991). They are used to describe instantaneous events. But in reality, people can't make distinction between a truly instantaneous event and an event taking really short time. Allen (1983) considered that most events could be decomposed into a series of smaller events. He gave an example for the event "finding the letter" that might be decomposed into "looking at spot X" and "realizing that it was the letter you were looking for". Another issue for these approaches is that if zero-width time point is allowed, then the question how time intervals (duration) can arise from zero-width time points (durationless) could not be answered (Allen 1983, Setzer 1991). A more practical approach is that a time point is viewed as a very small time interval and it may be decomposed into more tiny subparts (Allen 1983). Setzer (2001) provided a similar alternative that a time point has a finite duration but does not have a lower bound on the duration. It makes sense that time points can be building blocks of time intervals with the assumption of non-zero-width time points.

There are two ways to construct time intervals from points. Firstly an interval can be identified as an ordered pair of time points – t_1 and t_2 – where the interval is between t_1

and t_2. Secondly an interval can be identified as a fully ordered set of points of such that $\{t|t_1 < t < t_2\}$. Since time points are viewed as very small intervals, the later option is more advantageous for construction of other concepts (e.g. the definitions of 13 irreducible interval-interval relations from Allen 1983).

With temporal interval as primitives, new time interval can also be constructed from intervals. The new interval begins where the earlier one begins and ends where the later on ends.

2.2 Temporal Relations

Taking the second approach above that a time interval is formed by a fully ordered set of points of time, Allen (1983) defines five relations between intervals – *before, equals, overlaps, meet* and *during*. To provide a better computational model, the *during* relation can be divided into three relations – *during, starts and finishes.* Plus their inverses there are a total of thirteen relations between two intervals. The relationships can express any relationship that can be held between an ordered pair of intervals. The relations are shown in Figure 2.1.

Vilain (1982) defined five relations between time point and interval. These relations can express all possible relations that can be held between time point and time interval. The relations are shown in Figure 2.2.

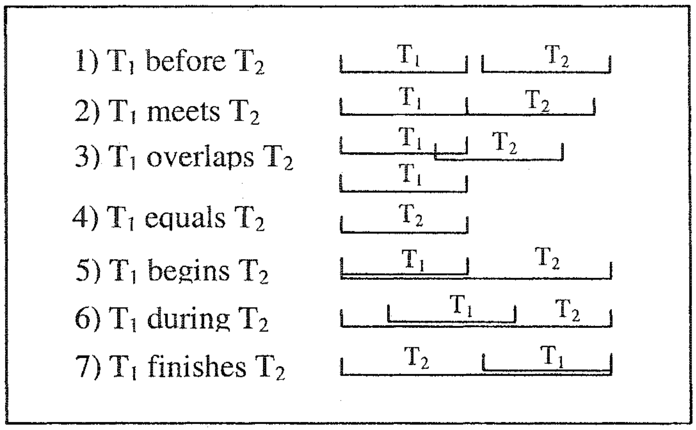

Figure 2.1: Relations between interval and interval (from Setzer 1991)

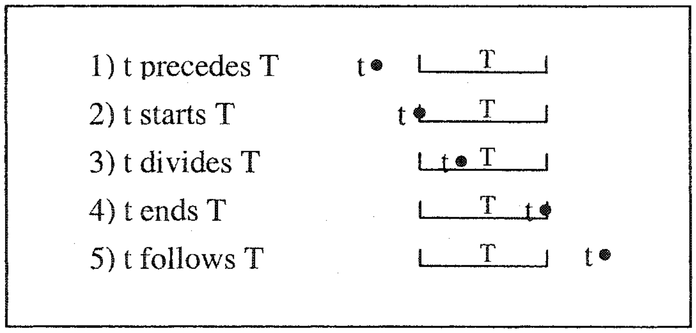

Figure 2.2: Relations between point and interval (from Setzer 1991)

The relations listed above are all the possibilities that can exist between two intervals or between a point and an interval. They are independent from the choice of instants or intervals as temporal primitives.

2.3 Representing Time

One of the most crucial problems in any computer system is to represent time. Many applications are facing this problem such as databases, simulation, expert systems and applications of Artificial Intelligence. Allen (1991) summarized some of the basic techniques available for representing time. Different representations are good for different situations about how precise the temporal information to be represented.

Approaches based on dating schemes:

Time stamp is in an absolute real-time format. For example, a convenient dating scheme could be a tuple consisting of the year, day in the year, hour in the day, minutes and seconds. The tuple (1990 110 10 4 50) could be interpreted as the 110th day of 1990, at 10:04(AM) and 50 seconds. Date-based representation is best for these applications where every event entered has its absolute date identified, e.g. a database maintaining transaction records on banking machines or a system recording patient's appointments for a doctor. The advantage of such a dating system is that time comparisons are reduced to simple numeric comparisons.

Constraint propagation approaches:

Constraint propagation approaches use graphs to represent times that are linked to each other with an arc labeled with the possible temporal relationships between the times. With this approach, new information added to a graph may constrain existing information because of transitivity constraints. Hence each time when adding new node to an existing graph, the arcs on the graph may have to be updated. Figure 2.3 is an example for a simple point-based constraint graph from Allen (1991).

This is a point-based representation. It has its limitation on representing arbitrary disjunction involving two or more points. For example, *e1* < *e2* or *e1* = *e2* can be expressed on an arc label (<,=) between nodes e1 and e2. But for an arbitrary disjunction such as *e1* < *e2* or *e2* < *e3*, there is no way to represent it on a graph. Allen (1983) developed an interval based constraint propagation algorithm that addressed this problem. For more information see Allen (1991).

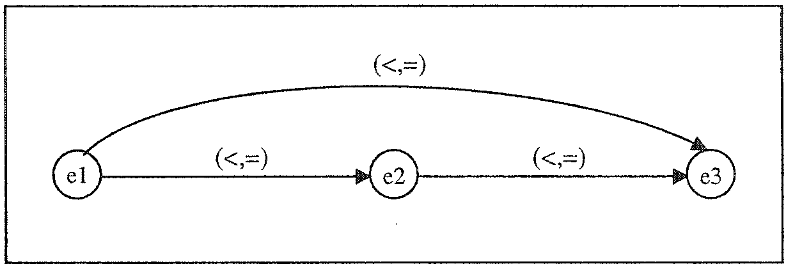

Figure 2.3: A simple point-based constraint graph

Duration-based representation:

There are two types of duration-based representation. PERT networks (Mehrotra et al. 1995) introduced a basic technique for dealing with durations of events. In a network, an acyclic directed graph that has distinguished beginning and ending event is used to maintain partially ordered events. Each node in the graph represents an event and has an associated duration. The numbers in a node indicate the earliest starting time and the latest starting time of an event. An arc labels duration of an event. However this representation is only useful in situations where the duration of each event is known. It becomes expensive when adding new events into the network because the network may have to be recalculated for the new information.

Another approach of duration-based representation uses duration between time points as base (Dean 1987). It encodes all temporal information in terms of duration constraints and uses heuristic graph search when constructing the graph. Each node of the graph represents a time point. The numbers on an arc indicate time units between two time points. This representation is good for situations in which duration information must be fairly well known for the most part. Figure 2.4 is an example from Allen (1991).

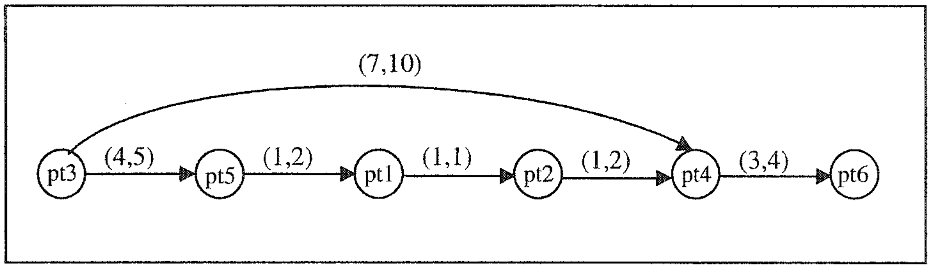

Figure 2.4: Time map using duration between time points as base

2.4 Temporal Annotation

During the last decade, remarkable progress has been made in the use of statistical techniques for analyzing text. These techniques demand large amount of annotated data for their development and testing. Many natural language processing applications such as information extraction, summarization, reasoning and question answering have benefited from these techniques. All these applications require robust extraction of events and temporal information to correctly position key events in time. This suggests that a solid annotation scheme is needed to annotate temporal information, events and relations between them.

Many research groups have devoted tremendous efforts to the development of temporal annotation schemes. Setzer (2001) introduced a conceptual framework and an annotation scheme (described in section 2.4.1). TIDES (Translingual Information Detection, Extraction, and Summarization) research program published the Temporal Annotation Guidelines (Ferro et al. 2001) (described in section 2.4.2). TERQAS Annotation Working Group released annotation guidelines for TimeML (Pustejovsky al et. 2002) in July of 2002 (described in section 2.4.3).

These annotation schemes enable events, times and temporal relations to be identified and marked up in text. For now these projects do not have the power to solve all the issues relating to temporal and event identification, e.g. implicit temporal expression, implicit temporal relation. However they do address some basic problem in the event-temporal identification such as time stamping events, ordering events, reasoning temporal expressions in context and reasoning events duration.

2.4.1 Setzer's Conceptual Framework

Setzer (2001) proposed a conceptual framework that can be used in classifying expressions in newswire texts. The framework presumes the world contains five primitive types: events, states, times, temporal relations and sub-event relations. The annotation scheme (STAG) turned from this framework is aimed at identifying events in newswire texts and to relate them to a calendar date or time on a time line. Thus the annotation scheme enables events, time and temporal relations and sub-events relations to be identified and marked-up. The annotation scheme does not annotate states because states are either not anchorable in time or are true over time span longer than that covered by the article.

2.4.1.1 Annotating Events

An event is something that happens at a given place and time. Some events can be viewed as conceptually instantaneous, others can be viewed as lasting over a certain period of time. In STAG, an event is treated as a black box without details. Thus the annotation scheme only annotates its representative, not the whole clause covering it. The criteria of whether something is to be annotated as an event is whether it is anchorable in time and can be placed on a time line.

In Setzer's annotation scheme, tag for annotating events is a *<event></event>* pair. There are a number of mandatory and optional attributes associated with events. The attributes for events are as following: *eid, class, argEvent, tense, aspect, relatedToEvent, eventRelType, relatedToTime, timeRelType, signalID.* Each event being annotated has an unique identifier *eid* and can be classified into groups - *class.* Attributes *tense* and *aspect* record information that is useful for determining when the event took place. The rest of the attributes are used to store information about relationship between the currently annotated event and other events, or between the currently annotated event and other time objects.

2.4.1.2 Annotating Time Expressions

Times can be viewed as intervals or time points. Setzer (2001) treat both of them equally as time objects that can be pinned down on a calendar date or time. Her annotation scheme applies to those time objects. Time expressions like *recently, few days ago* that are not possible to associate to a calendar date or time are not considered as time object. Thus they are excluded from the annotation scheme.

The tag for annotating time is a *<timex></timex>* pair. Three non-optional attributes are associated with the tag, *tid, type, calDate*. Like events, each time object being annotated has an unique identifier – *tid*. Attribute *type*, whose possible values are DATE and TIME, is used to indicate whether the unit of a time object is longer than a day. Attribute *calDate* is the calendrical date/time represented by the time object. One of the format for *calDate* is [[DD]MM]YYYY, which example value "12031996" means *March 12, 1996*.

2.4.1.3 Annotating Temporal Relations

Events in newswire usually are related to other events in certain ways. They may also be related to other time expressions. In STAG, there is no specific tag for annotating temporal relations. Rather the relationships between two events, or between one event and other time objects are stored in an event tag. The attributes *relatedToEvent* and *eventRelType* serve the purpose of storing relationships between events, and the attributes *relatedToTime* and *timeRelType* are used to store relationships between events and other time expressions. Which event has to carry relationship information depends on the type of relation.

Some words (usually conjunctives or propositions such as *when, while, after, on*) serve as signals to indicate relations between events or events and times. Such words are called signals. A signal plays a very important role in annotating temporal relations so that there is a specified tag designed for it. The tag for signal is a *<signal></signal>* pair. If a signal appeared in a text, then not only is the signal annotated but also the *ID* of the signal is stored in *signal* attribute, so the link between the signaling word and the events is not lost. For some implicit temporal relations, signals are not presented between related events and/or times.

2.4.1.4 Examples Using STAG

Two examples from Setzer (1991) illustrate how the annotation scheme is used to annotate events, times and temporal relations.

The example shows the annotation of temporal relation between two events:

```
All 75 people on board the Aeroflot Airbus
<event eid=4 class=OCCURRENCE tense=past relatedToEvent=5
        eventRelType=simultaneous signalID=7> died </event>
<signal sid=7> when </signal> it
<event eid=5 class= OCCURRENCE tense=past> ploughed </event>
into a Siberian mountain.
```

Another example shows the annotation of temporal relation between event and time object:

```
A small single-engine plane
<event eid=9 class=OCCURRENCE tense=past relatedToTime=5
        timeRelType=is_included signalID=9> crashed </event>
into the Atlantic Ocean about eight miles off New Jersey
<signal sid=9> on </signal>
<timex tid=5> Wednesday </timex>.
```

2.4.2 TIDES Temporal Annotation Guidelines

TIDES (Translingual Information Detection, Extraction and Summarization) published a document – TIDES Temporal Annotation Guidelines - for annotating time expressions with a canonical representation of the times they refer to (Ferro et al. 2001). These guidelines intend to support a variety of downstream applications in the performance of some useful tasks such as information extraction, question answering and summarization.

In the TIDES annotation scheme, temporal expressions are treated as stand-alone targets for annotation and extraction so that the semantic representation allows the scheme to be highly language-independent. The annotation scheme uses time points as primitive, which is aimed at using the point representation to the extent possible. This enforces consistency in tagging temporal expressions among different annotators.

The TIDES annotation scheme is designed to meet certain criteria (Ferro et al. 2001):

Simplicity with precision: The annotation scheme should be simple enough for humans to execute and precise enough for use in various natural language processing tasks.

Naturalness: The annotation scheme should reflect those distinctions that a human could be expected to reliably annotate.

Expressiveness: Time value used in annotation scheme should be as full as possible, within the bounds of what can be confidently inferred by annotators.

Reproducibility: The annotation scheme should be as little ambiguous as possible to ensure consistency among annotators.

2.4.2.1 Annotating Time Expressions

The TIDES annotation process consists of two major steps: flagging temporal expressions, and identifying corresponding time values for the temporal expressions. The annotation scheme addresses mainly three different kinds of time values: time points (answering "when?" questions), durations (answering "how long?" questions), and frequencies (answering "how often?" questions). However there are several semantic problems of temporal expressions such as fuzzy boundaries and non-specificity that are also addressed.

Like STAG, TIDES's temporal annotation scheme is not attempting to flag every time related expression. It follows two basic principles in applying temporal annotation scheme (Ferro et al. 2001). Firstly, the time value of the temporal expression must be able to be determined by a human. Secondly, the time value of the temporal expression must be based on evidence internal to the document that is being annotated. Both local and global context can be used to determine the time value of a temporal expression. Thus temporal expressions like *recently* and *lately* are not considered as markable temporal expressions and will not get annotated in the document.

The tag used by TIDES for annotating temporal expressions is *<timex2></timex2>* pair. There are a number of attributes associated with the tag element. *VAL* is the most important attribute of all. It is the time value represented by the annotated temporal expression. It can be used for any temporal expression that can be pinned down to a time point (or interval) on a calendar/clock. The format of the value of *VAL* is based on the formats defined by ISO 8601. Example values are "1999-10-01" for October 1, 1999, and "1999-10-01-T09" for 9 a.m. Friday, October 1, 1999. In case that the value of any position is unknown, a placeholder character, X, is used in that position.

Other attributes are *MOD, SET, PERIODICITY, GRANULARITY, NON_SPECIFIC* and *COMMENT*. These attributes are not to be described in details here. For detailed information, consult "TIDES Temporal Annotation Guidelines" (Ferro et al. 2001).

2.4.2.2 Examples Using the TIDES Annotation Scheme

For the sake of reproducibility, Ferro (2001) provided an example-based approach to ensure consistency among annotators. Each guideline is closely tied to specific examples. Here are some examples from the document of TIDES Annotation Guidelines to demonstrate how the annotation scheme is applied to annotate temporal expressions. The time values *VAL* of temporal expressions in examples (from Ferro 2001) are determined either locally or globally.

The following examples show the annotation of anchored expressions, dates and times:

The bombing took place on

<TIMEX2 VAL="1998-12-02"> the second of December </TIMEX2>.

The sponsor arrived at

<TIMEX2 VAL="1999-07-15-T14:50"> ten minutes to 3 </TIMEX2>.

The following example shows the annotation of duration:

The video is only

<TIMEX2 VAL="PT30M"> half an hour long </TIMEX2>.

Time range is annotated by TIDES as two separate time points. Below is an example that shows the annotation of explicitly defined ranges:

The class is

<TIMEX2 VAL="1999-07-15-T15"> 3 </TIMEX2>

to

<TIMEX2 VAL="1999-07-15-T18"> 6 pm today </TIMEX2>

The TIDES annotation scheme allows embedded tags used for complex temporal expressions. Below is an example that shows the annotation of temporal expression using embedded tag:

Saddam might play the whole game again

<TIMEX2 VAL="2000-01"> six months </TIMEX2>

or

<TIMEX2 VAL="2000"> a year from

<TIMEX2 VAL="PRESENT_REF"> now </TIMEX2></TIMEX2>.

2.4.3 TimeML Annotation Guidelines

TERQAS (Time and Event Recognition for Question Answering Systems) was an annotation working-group whose purpose was to address the problem of how to answer temporal-based questions about the events and entities in news text. The workshop has delivered its research result, TimeML Annotation Guidelines (Pustejovsky et al. 2002), an annotation scheme for annotating events and time entities. The annotation scheme differs from most previous attempts at event and temporal specification in that TimeML separates the representation of event and temporal expression from the anchoring or ordering dependencies that may exist in a given text.

TimeML is modeled on both STAG (Setzer 2001) and TIDES (Ferro et al. 2001). Its annotation guidelines for events are close to STAG's, and its annotation guidelines for temporal expressions are close to TIDES's. Tags and their attributes from both STAG and TIDES are seen in TimeML.

2.4.3.1 Annotating Events

As in Setzer's conceptual network, TimeML considers an event as something that happens now or will happen soon. Events can be punctual or last for a period of time. In news text, events can be expressed by means of verbs, nominalizations, adjectives, predicative clauses or prepositional phrases (Pustejovsky et al. 2002).

The tag for annotating event is *<event></event>* pair. The attributes, *eid, class, tense, aspect,* are similar to those in Setzer's framework (some differences exist in specification). Unlike Setzer's, temporal relations between events, or events and time are not encoded in event tag. Instead, various temporal relations are recorded by using a set of LINKs.

2.4.3.2 Annotating Time Expressions

TimeML provides guidelines for marking up explicit temporal expressions including dates, times and durations. Temporal expressions maybe fully specified (e.g. March 19, 2004) or underspecified (e.g. last year, Monday). The tag used to annotate` time expressions is *<TIMEX3></TIMEX3>* pair. Some of the attributes from both TIMEX and TIMEX2 are kept in TIMEX3.

2.4.3.3 Annotating Temporal Relations

Setzer's annotating scheme (STAG) encodes temporal relations into attributes of *<event>* tag. TimeML uses a different approach to express temporal relations. A set of

LINKs and SIGNALs (first introduced by Setzer (2001)) are used to encode various relations between the temporal elements of a document.

- TLINK (Temporal Link) is used to represent the temporal relationship between events or between an event and a time.
- SLINK (Subordination Link) is used for context introducing relations between two events, or an event and a signal.
- ALINK (Aspectual Link) is used to represents the relation between an aspectual event and its argument event.
- SIGNAL is used to indicate how temporal objects are to be related to each other. Temporal prepositions *(on, during*), temporal connectives (*while, when*) and subordinators (*if*) are among the candidates for SIGNAL.

2.4.3.4 Examples Using TimeML

An example (from Pustejovsky et al. 2002) is used here to illustrate how to use TimeML to annotating events and temporal expression.

```
John
<EVENT eid="e1" class="OCCURRENCE" tense="PAST"
aspect="PERFECTIVE"> left </EVENT>
<MAKEINSTANCE eiid="ei1" eventID="ei"/>
<TIMEX3 tid="t1" type="DURATION" value="P2D" temporalFunction="false">
                2 days </TIMEX3>
<SIGNAL sid="s1">before</SIGNAL>
the
```

```
<EVENT eid=”e2” class=”OCCURRENCE” tense=”NONE”
        aspect=”NONE”> attack </EVENT>
<MAKEINSTANCE eiid=”ei2” eventID=”e2”/>.

<TLINK eventInstanceID=”ei1” signalID=”s1” relatedToEvent=”ei2
relType=”BEFORE” magnitude=”t1”>
```

2.4.4 Differences Among Annotation Schemes

TimeML is considered the most complete temporal annotation scheme so far. It is modeled on both Setzer's work (Setzer 2001) and the TIDES's work (Ferro et al. 2001) (indeed Setzer and Ferro are both members of TERQAS Annotation Working Group). The working group utilized ideas from Setzer's event annotation, temporal relation annotation and TIDES's temporal expression annotation and produced a better, more complete scheme. There exist features in one annotation scheme that do not exist in others. Table 2.1 summarizes the feature differences among the annotation schemes.

	STAG	**TIDES**	**TimeML**
Event	+	-	+
State	-	-	+
Date	+	+	+
Time	+	+	+
Duration	-	+	+
Temporal relation	+	-	+
Signal	+	-	+
Embedded tags	-	+	-

Table 2.1: Differences among annotation schemes

2.5 Discourse Structure Analysis

There is only one real-world event structure, but different authors may organize events into many different discourse structures. The frameworks in discourse structure analysis provide mechanism for extracting temporal information from text.

2.5.1 Personal Narratives

Labov and Waletzky (1967) developed a framework for analysis of narratives of personal experience told in conversation. The framework focuses mainly on six elements: abstract, orientation, action, evaluation, resolution and coda. These six elements occur in the order as listed in personal narratives. Because of the differences between genres, the elements and their ordering maybe vary among frameworks that are developed for different genres.

2.5.2 News Stories

Bell introduced a framework for analyzing the discourse structure of news (Bell 1995a; Bell & Garrett 1997). The framework focuses on the question "what does this story actually say happened?". It provides guidelines for building structures for four aspects: events, time, place and news actors. The elements that are needed to describe the discourse structure of news stories are listed below.

Attribution: includes credit to news agency, journalist's byline, setting of place and time.

Abstract: consists of lead sentence that covers the central event of the story and possibly one or more secondary events.

Story: consists of episodes that describe one or more events. Events describe actors, action, and setting of time and place.

Background: covers and events prior to the current action – story past time.

Commentary: provides the journalist's or news actor's observations on the action, assessing and commenting on events as they happen – present time.

Follow-up: covers any action subsequent to the main action of an event – story future time.

In Bell's approach, the discourse structure of news stories can be organized in a tree structure. Elements described above form nodes of the structure tree. Figure 2.5 is an example of discourse structure of a single-sentence story from Bell and Garrett (1997).

Clashes kill eight

At least eight people have died in tribal fighting in the Bimbila region of northern Ghana. – AFP

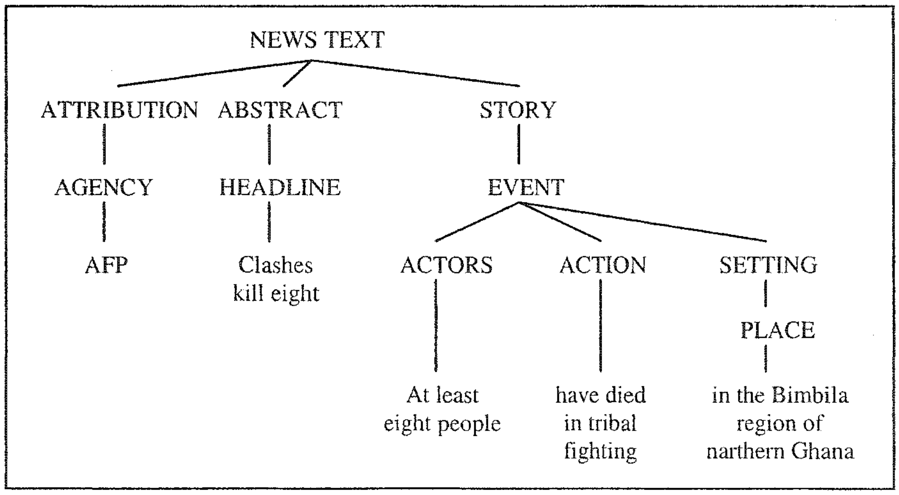

Figure 2.5: Discourse structure of news story

The time structures of news events by using this framework are constructed on sentence basis. Every sentence is assigned an integer value indicating the order in which the events in the sentence happened. Table 2.2 shows how a time structure is built.

Sentence #		Time Structure
S1	Lima, Jan 18. The estranged wife of Peru's President Alberto Fujimori was taken to hospital today just 24 hours after she began a hunger strike to protest at her party's elimination from congressional elections.	0 -2 -3
S2	Doctors said she was suffering from tachycardia, or an accelerated heartbeat.	+1
S3	Earlier, [deposed first lady Susanna] Higuchi, sitting under an umbrella in a scorching summer sun outside the National Electoral Board's headquarters, had pledged to press on with her protest.	-1
S4	The electoral board said on Monday Higuchi's Armonia-Frempol party had not qualified for the April Congressional vote because it failed to present a full list of candidates for the 120-member legislature.	-3 +2 -4
S5	Board member Manuel Catacora said today that since Higuchi had presented her party's congressional slate just 10 minutes before the filing deadline, a provision allowing parties five days to correct and error did not apply.	0 -4 0
S6	Higuchi, a 44-year-old civil engineer, has been estranged from Fujimori since August when she protested an election law that banned her from running from [sic] public office.	-5 -6

Table 2.2: Time structure based on sentence

CHAPTER 3

TEMPORAL STRUCTURE OF NEWS STORIES

A news story covers information about something that has just happened or will soon happen. They are the daily happenings of our societies expressed in the media. In this project we target news stories to analyze the temporal structure. Our goal is to find out any possible structure pattern that journalists use to write their news stories.

Since text is central to news, analysis of news stories is really analysis of the news texts. Some research on text analysis focus on sentences. Our study is based on paragraphs. By determining a story time for each paragraph we derive a temporal structure of the news story.

All the data we used for our analysis are news stories obtained from APW website for the DUC-2002 meeting. All the documents are in different format. There are twelve clusters of news stories that have been studied. The news stories cover various topics, e.g. economics, politics, disasters and science, etc.

3.1 Annotation Schemes

As discussed in section 2.4.4, TimeML is considered the most complete temporal annotation scheme. It provides guidelines for annotating events, time expression and temporal expression. Before modeling the time structure for a news story, we have to have the knowledge about story time for each paragraph. We achieve this task by annotating each of the paragraphs in the news text. The time value of a paragraph is determined either by temporal expressions in the paragraph or by the context. Both scenarios require knowing the real time value for temporal expressions in the current paragraph or previous paragraph. This is done by manually annotating the temporal expressions in text.

TimeML can be used to annotate temporal expressions in news text. However our task is only to know the value of the temporal expressions. Our annotation scheme is based on TimeML (Pustejovsky et al. 2002) but with some modifications. We will describe the details latter.

3.1.1 Paragraph

A paragraph is a subdivision of a text intended to separate ideas. Each paragraph is a block of text. To acquire time structure of a news story, we annotate all the paragraphs in its text no matter how long or how short to get the story time for the paragraph.

3.1.1.1 Annotation of Paragraphs

The tag for paragraph is <PARAGRAPH> which must be closed by a closing tag </PARAGRAPH>. One paragraph can only apply one paragraph tag. Below is an example use of <PARAGRAPH> tag.

<PARAGRAPH ID="2" VALUE="0">

> *The long-awaited opening of the restaurant on one of Belgrade's main downtown squares will take place March 24, the Yugoslav news agency Tanjug reported, and it will offer Big Macs, fries and the other specialties familiar to McDonald's customers in the West.*

</PARAGRAPH>

3.1.1.2 Attributes for PARAGRAPH Tag

a. Paragraph identification number (ID)

ID = {integer}.

ID is a non-optional attribute. Each paragraph has to be identified by a unique ID number. The ID of the first paragraph is 1 (one). It increment by 1 (one) every time we assign an <PARAGRAPH> tag to a paragraph.

b. Story time value of paragraph (VALUE)

VALUE={integer}

VALUE is a non-optional attribute. The VALUE of a paragraph is either an integer or a set of integer depending on story times of actions in that paragraph.

3.1.2 Temporal Expressions

A temporal expression is defined as a chunk of text that expresses some sort of direct or inferred temporal information. It represents the real time point or duration in which something happened or will happen in the news story. All the real time values for temporal expressions must be based on evidence internal to the news text that is being annotated.

3.1.2.1 Annotation of Temporal Expressions

The presentation of time values we use in this project differs from TimeML in terms of their format. TimeML uses the formats defined by ISO 8601 that includes days of week. Absolute dating system is used for representing time point and duration in our study. This involves time stamping for each markable temporal expression with an absolute real-time value. But we use a different format from Allen's (yyyy ddd hh mm ss) (Allen 1991). The way we used to represent time is a string of "yyyy-mm-dd-hh-nn-ss" where y, m, d, h, n, s are all integers. This is also a standard format defined by ISO 8600. Although the exact time point for a temporal expression may not be able to be identified in some situations, we still can keep this format by applying the idea of placeholder to fill the position that is unknown.

The big advantage of this time representation is that it provides an algorithm for comparing times involving only number comparison if values on all the positions are known. Even if placeholders are used in some of the positions, we can still compare the numbers in different slots. This also provides some precision on the comparison. For example, we can't compare "1989-09-18-xx-xx-xx" and "1989-09-xx-xx-xx" to determine which one occurs first, but we can certainly compare "1989-09-18-xx-xx-xx" and "1989-11-xx-xx-xx-xx".

The tag for temporal expression is <TIMEML> which must be closed by a closing tag </TIMEML>. Below is an example use of TIMEML tag.

> *The communist world gets its first McDonald's*
>
> *<TIMEML ID="1" TYPE="date" VALUE="1988-03-20-xx-xx-xx" VALUE2="1988-03-26-xx-xx-xx"> next week </TIMEML>,*
>
> *and some people here are wondering whether its American hamburgers will be as popular as the local fast-food treat, Pljeskavica.*

3.1.2.2 Attributes for TIMEML Tag

a. Temporal expression identification number (ID)

ID = {integer}.

ID is a non-optional attribute. Each temporal expression has to be identified by a unique ID number. The ID of the first temporal expression is 1 (one). Incremented by 1 (one) every time we assign an TIMEML tag to a temporal expression.

b. Type of temporal expression (TYPE)

TYPE = {date, duration}

TYPE is a non-optional attribute. Every 'markable' temporal expression must have a TYPE to indicate the property of real time inferred by the temporal expression.

c. Start point of temporal expression in real time (VALUE)

VALUE = {yyyy-mm-dd-hh-nn-ss} where y, m, d, h, n, s are all single digit positive integers. For a temporal expression, if any position of y, m, d, h, m, s is unknown to annotators, then the position is replaced by a placeholder x.

VALUE is a non-optional attribute. Every 'markable' temporal expression must have a VALUE. It represents a time point or a start point of a time interval inferred by the temporal expression.

d. Ending point of temporal expression in real time (VALUE2)

VALUE2 = {yyyy-mm-dd-hh-nn-ss} where y, m, d, h, n, s are all single digit positive integers. Similarly to the VALUE attribute, placeholder x may be used at any position that is unknown to annotators.

VALUE2 is an optional attribute. It applies only to time intervals with identifiable ending points. It represents an end point of a time interval inferred from the temporal expression.

3.1.2.3 Special Cases on Attributes VALUE and VALUE2

VALUE and VALUE2 are the attributes that represent real time point or real time duration of the annotated temporal expression. Since time is eternal, there is expression limitation on the format of these attributes. For some temporal expressions these attributes may not have the power to express the real time. For example,

> *The universe was created by a tremendous explosion about 15 billion years ago.*

The temporal expression - 15 billion years ago - is not expressible by using the format "yyyy-mm-dd-hh-nn-ss". Therefore two special values for VALUE and VALUE2 are introduced to accommodate this kind situation. Although ambiguities may still exist in those special values, they are sufficient for this particular project.

0000-00-00-00-00-00 is used when the temporal expression represents a time point or duration that is long long time ago and out of the scope of the format. On the opposite side, 9999-99-99-99-99-99 is used when the temporal expression represents a time point or duration that is in future and out of the scope of the format.

3.1.2.4 Markable Temporal Expressions

There are two types of temporal expressions that may qualify as "markable" temporal expressions, dates and durations.

a. Date: The expression describes an absolute/calendar time or a time of the day. All temporal expressions which belong to type *date* are markable.

Examples: *Mr. Brown left* ***Monday, March 24, 1988***
on ***March 24, 1988***
in ***March, 1988***
in ***the summer of 1988***
on ***Sunday***
at ***ten in the morning***
the morning of March 24
last night

b. Duration: The expression describes a time interval.

Examples: *Mr. Brown stayed* ***2 months*** *in Montreal.*
three weeks
10 days in March
four hours yesterday

For some date ranges such as "from 5 till 6", they are not considered as interval. Two separate <TimeML> tags are used when they are annotated. Propositions are not included in tags.

3.1.2.5 Non-markable Temporal Expressions

Not all temporal expressions are markable. Only these explicit times and/or durations like the above examples can be annotated. Temporal expressions that can not be pinned down on a calendar as time point or duration are considered as non-markable temporal expressions. The VALUE attribute is unknown to the annotator when one attempts to annotate such temporal expressions.

Here are examples of non-markable temporal expression as time:

> *Mr. Brown left* ***few days ago.***
> *Mr. Brown left* ***recently.***

The temporal expressions "few days ago" and "recently" can 't be identified as exact time points on a calendar or clock so they are non-markable temporal expressions.

Here is an example of non-markable temporal expression as duration:

> *Mr. Brown has stayed in Montreal for* ***years.***

The temporal expression "years" doesn't have a clear-cut edge on either end as duration. We can't figure out exactly how long Mr. Brown has stayed, when he arrived in Montreal. It is a non-markable temporal expression as duration.

3.1.3 Guidelines of Annotating Paragraphs

Time structure of a news story is the order in which events are told. Modern news stories are usually non-chronological so that the order the events being told is usually not the same as the order events happened.

The Bell's framework (Bell 1995a) for analyzing news discourse structure uses integer number to indicate the order of events happening. The values are assigned to each sentence according to the order of actions in the sentence happened. The time structure is built on those values.

This project is to study patterns of time structure of news stories. Our approach is based on the paragraphs, not the sentences. When analyzing text, we assign each paragraph a story time. Since a paragraph consists of sentences that usually describes the same event, or events that are very closely related, most paragraphs have only a value of one integer in story time. However there are many scenarios where a paragraph may have multiple values.

We developed some guidelines for assigning story time for paragraphs:

1. If only one action exists in a paragraph, the paragraph is assigned one value only. The value of the paragraph depends on the action.

```
<PARAGRAPH ID="15" VALUE="0">
    The officials urged residents in the higher risk areas along the south coast
    to seek higher ground.
</PARAGRAPH>
```

2. Two or more actions exist in a paragraph, if the actions virtually occurred at same time or on immediately after another, the paragraph is assigned one value only. The value of the paragraph depends on the actions.

<PARAGRAPH ID=”1” VALUE=”0”>

> *A major earthquake rocked northern California Tuesday evening, collapsing part of the San Francisco Bay Bridge and shaking Candlestick Park and buildings up to 95 miles away.*

</PARAGRAPH>

3. A paragraph just provides general information on something. There is no time bonded to the object being described. The information may be true at all the time. We assign such paragraphs a story time as same as story present time, Time zero.

<PARAGRAPH ID=”4” VALUE=”0”>

> *Pljeskavica is made of ground pork and onions, and it is served on bread and eaten with the hands. It is sold at fast-food restaurants across the country and costs about a dollar.*

</PARAGRAPH>

4. There are more than one temporal expressions in one paragraph. If two or more temporal expressions have different time points or durations on the calendar or clock, the VALUE attribute of a paragraph may have multiple values.

<PARAGRAPH ID=”20” VALUE=”-6,-5”>

> *Honecker was ousted from the leadership on Oct. 18, after massive pro-democracy demonstrations. He was expelled from the communist Party in December and has since been accused of treason and abusing his power.*

</PARAGRAPH>

5. If there are two or more actions that do not occur at the same time or one occurs immediately after another, the VALUE attribute of a paragraph may have multiple values.

> *<PARAGRAPH ID="0" VALUE="0,-1">*
>
> > *The communist world gets its first McDonald's next week, and some people here are wondering whether its American hamburgers will be as popular as the local fast-food treat, Pljeskavica*
>
> *</PARAGRAPH>*

6. A paragraph cites other media's or other people's reaction, comments. We only focus on what the other parties are saying not on who is saying. The story time of the paragraph depends on what is cited.

> *<PARAGRAPH ID="5" VALUE="0">*
>
> > *``In fact, this is a clash between the Big Mac and Pljeskavica," said an official of Genex, Yugoslavia's largest state-run enterprise that will operate the McDonald's.*
>
> *</PARAGRAPH>*

3.2 Process News Story

Before we start annotating news text, we are going to clarify a very important fact, the date of the news story published. Because of the characteristic of news, many temporal expressions in a news story depend on that date to determine their time values.

3.2.1 Time Stamp of News Story

"Technological development in the pursuit of timeliness continues to impel news coverage towards 'present-acion' – closing the gap between the event and its telling, with the goal of displaying events in 'real time'." (Bell 1995a)

News stories are perishable in time. Immediacy is a very important characteristic of news. Both journalists and readers presume that the new story is something which has just happened. Journalists usually do not use full time expressions to indicate the story time. They most likely tell reader that the story happened "yesterday" or "today". Sometimes there is no overt time reference at all in news stories. So the news story time stamp is very crucial for the computation of the value of most temporal expression. It serves as index for the temporal expressions. For example 'yesterday', it is impossible to know what day is yesterday without knowing the news story time stamp. Even for most explicit temporal expression, e.g. *Oct. 18*, it can only be evaluated with respect to the year of date that the news story was written.

In the data we analyzed, time stamp of news story is recorded in *<DOCNO>* and *<FILEID>* field of the news text. It indicates the date when the news story is published.

3.2.2 Steps to Process News Story

The following steps were performed to analyze a news story to get the time structure.

a. Locate all the temporal expressions in news text.

b. Identify story present time.

c. Identify story time for every temporal expression.

d. Annotate temporal expressions.

e. Build up time structure of news text in terms of paragraphs.

f. Annotate paragraphs.

g. Draw time structure chart.

h. Once all the target texts are analyzed and annotated, we can analyze time structures to find any possible patterns.

3.2.3 Analyzing News Texts

We now use a news story titled “First McDonald's to Open in Communist Country” to illustrate how the analysis is done. The story is a typical international news agency story published by the American Press, March 14, 1998. It is about America’s fast food chain giant McDonald’s and its first restaurant in a communist country. It is short and simple but enough for the task of demonstration.

3.2.3.1 Locate Temporal Expressions

For now we locate temporal expressions manually by reading the story and highlighting all temporal expressions, including markable temporal expressions and non-markable temporal expressions. At some point we may use some existing system to identify temporal expressions. All the temporal expressions are used in determining story times for paragraphs, but only these markable temporal expressions are annotated in latter step.

3.2.3.2 Identify Story Present Time

The second step of analysis is to find out what is the story present time. In modern news writing, all the journalists' basic facts are usually concentrated at the beginning, then expanded further down. So most likely we can find when, the story present time, in the first few sentences.

In most cases, the first sentence of the news story summarizes the central action and establishes the point of the story. This sentence is called 'lead' and it serves the function of an abstract. It is what the journalist wants to tell the reader: what exactly happened and when it happened. The time when the event in the 'lead' actually happened is defined as story present – Time zero. All other story times use Time zero as an anchor to determine their values.

In our example, the first sentence is also the first paragraph of the story. It does not give us an explicit date and/or time when the restaurant opened, because the only temporal expression in the first sentence is "next week". This temporal expression has meaning only if we know the day when the news story is written. According to the *<DOCNO>* or *<FILEID>* of the text, we can know this news story was published on March 14, 1988. If we look at the calendar, the "next week" would be from March 20 to March 26, 1988. Obviously this range is too broad to be story present since most of the events in a news story would happen around story present time – the time the main event happened. A wide range story present makes it difficult to determine the story time for each of the paragraphs. Most events in the same news story may fall into the same week.

More precise story present time of the example text is given in the second paragraph. The temporal expression “March 24” tells us exactly when the restaurant opening will take place. We define *March 24* as Time zero of the story.

3.2.3.3 Identify Other Time Points in News Story

Now we have explicitly identified story Time zero. There may be other earlier or later time points that appear in later sentences. In order to build up a time structure of the news story we need to know story times for every paragraph. Temporal expressions are the key to determine story times. We know when and what event happened in news story according to these temporal expressions.

Time references in news stories are expressed in various ways. Some are expressed in absolute/calendar time, e.g. “March 24”. Some are expressed in relation to other time points, e.g. “for years”. Others are deictics (“today”) with the present as reference point.

Here we pick up temporal expressions in all formats from text. Using Time zero as anchor, we determine values of each temporal expression according to how far it is from Time zero. This is a partial order. Time points prior Time zero on calendar will have negative values. In contrast, time points after Time zero will have positive values. The further the time point from Time zero, the bigger the value is. The result is listed in Table 3.1.

Temporal expression	Paragraph	Calendar time	Story time
next week	1	> March 20, 1988 && < March 26, 1998	0
March 24	2	= March 24, 1988	0
by the end of this year	7	> March 24, 1988 && < December 31 1988	+1
for years	8	< March 18, 1988	-2

Table 3.1: Temporal expressions in McDonald's opening story

The first two temporal expressions in Table 3.1 really mean to be the same time point when McDonald's restaurant will open. "March 24" is just more precise expression for this time point. They all represent story present time. The main event in this particular news story will happen at this point so that they are all defined as Time zero.

The seventh paragraph is the only one that describes an event that occurs after the main event. That is indicated by the temporal expression "by the end of this year". The story time of this temporal expression must have a positive value. No other events happen before it and after the Time zero so that the story time of this temporal expression is +1.

The temporal expression "for years" in the last paragraph tells us that the events happened in this period of time is before the time point when main event occurred. A negative value for this time is ensured. Is there any event that happened between it and Time zero? When we read this news story we can find in third paragraph that media have given suggestions on the success of the American restaurant. Although the story didn't explicitly indicate when, it is reasonable to assume that the suggestions come after the negotiations are done and before the restaurant opening. So the story time given to this temporal expression is –2. –1 is reserved for the media suggestion that is not listed on the Table 3.1.

3.2.3.4 Annotating Temporal Expressions

Although this project studies time structures of news stories in terms of paragraphs, we annotate not only paragraphs but all temporal expressions that can be annotated as defined before. The annotated texts maybe used by other members in our group for their research.

As the definition of VALUE and VALUE2, the real time beginning and ending are expressed as the format of "yyyy-mm-dd-hh-nn-ss" where y, m, d, h, m, s are all single digit integers if we know actual date and time of the temporal expressions. Otherwise a placeholder "x" is used in corresponding position. A duration has different real time beginning and ending while a time point has the same real time beginning and ending. All temporal expression and their real calendar/clock time are listed in Table 3.2.

Temporal expression	Type	Real time beginning	Real time ending
next week	Date	1988-03-20-xx-xx-xx	1988-03-26-xx-xx-xx
March 24	Date	1988-03-24-xx-xx-xx	1988-03-24-xx-xx-xx
by the end of this year	Date	1988-03-24-xx-xx-xx	1988-12-31-xx-xx-xx
for years	Duration	N/A	1988-03-18-xx-xx-xx

Table 3.2: Real time represented by temporal expressions in McDonald story

The temporal expressions "next week" and "by the end of this year" are both presenting a period of time. But since the restaurant opening event only happens in a particular time in the time period, according to the definition of TYPE, their types belong to date, not duration. The last temporal expression "for years" is a type of duration because negotiations go on and off for a long time. Since we can't figure out when first meeting started, the real time beginning is not applicable. For now we don't annotate this temporal expression.

3.2.3.5 Building the Time Structure Based on Paragraph

Based on story time of the temporal expressions, we build the story's time structure as listed in Table 3.3.

In the first paragraphs, we can find two actions. One is "the communist world get its first McDonald's" and another is "some people here are wondering". They are related events but they occurred at different time points that are apart for few days. These two actions happened at different story times so that it would not be accurate if we only give one time point for such two actions in the story time structure. The first action is the lead action that happened at Time zero. The second action happened before the lead action and it seems there are no other events between it and the lead action. Its story time has a value of –1.

	P#	Time structure
The communist world gets its first McDonald's **next week**, and some people here are wondering whether its American hamburgers will be as popular as the local fast-food treat, Pljeskavica.	1	0, -1
The long-awaited opening of the restaurant on one of Belgrade's main downtown squares will take place **March 24**, the Yugoslav news agency Tanjug reported, and it will offer Big Macs, fries and the other specialties familiar to McDonald's customers in the West.	2	0
The Belgrade media have suggested that the success of the American restaurant depends on its acceptance by Yugoslavians who are long accustomed to the hamburger-like Pljeskavica.	3	-1
Pljeskavica is made of ground pork and onions, and it is served on bread and eaten with the hands. It is sold at fast-food restaurants across the country and costs about a dollar.	4	0
``In fact, this is a clash between the Big Mac and Pljeskavica,'' said an official of Genex, Yugoslavia's largest state-run enterprise that will operate the McDonald's.	5	0
John Onoda, a spokesman at McDonald's Oak Brook, Ill., headquarters, said it was the first of the chain's outlets in a communist country.	6	0
The next East European McDonald's is scheduled to be opened in Budapest, Hungary, **by the end of this year**, said Vesna Milosevic, another Genex official.	7	+1
Negotiations have been going on **for years** for expanding the fast-food chain to the Soviet Union, but no agreement has been announced.	8	-2

Table 3.3: Time structure in McDonald's opening story

Journalists often use reported speech to describe or comment on what happened. It tells readers the source of this piece of information and may add some credibility to the news story. It does not affect the fact that is already happened no matter whoever is saying. News consumers may only care about what happened and ignore who provided this information to them. The second paragraph illustrates this situation. The author of this news story cites something the Yugoslav news agency Tanjug reported. The main point of this paragraph is what Tanjug reported, not who reported it. The action "opening will take place" is the main event of this story at Time zero.

The third paragraph, the story is moving back up to Time -1, the Belgrade media have some suggestions prior the restaurant's opening. It happened in the same time period as the second action in the first paragraph. They have some relationship but the reader can't figure out which action is performed first. Although they are different actions they share the same story time.

The forth paragraph is a general description of Pljeskavica, a kind of food served in local restaurant. It is true for all the time so that we treat it as story present time, Time zero.

Paragraphs five and six are similar to the second paragraph. They cite other parties' comments on the main event.

The seventh paragraph is the only one that describes action in future from the main event. The action "the next East European McDonald's is scheduled to be opened in Budapest" is few months later after Belgrade's opening. It has a story time of the smallest positive value after Time zero.

The last paragraph tells readers that before the restaurant open there are negotiations going on. Although we don't know when the first negotiation exactly began, "for years" presents the earliest time point in this news story. So this paragraph owes the earliest story time –2.

3.2.3.6 Example of Annotated Text

The text below is the example after annotation. More annotated texts are listed in appendix A. Since our annotation schema is compliant with XML notation, the annotated text is still well formatted. There are three temporal expressions in this example that can be annotated. One is "next week" in the first paragraph, another one is "March 24" in the second paragraph, other one is "by the end of year" in seventh paragraph. So there are three TIMEML tags added when we annotate this news story.

```
<PARAGRAPH ID="1" VALUE="0, -1">
    The communist world gets its first McDonald's <TIMEML ID="1" TYPE="date"
    VALUE="1988-03-20-xx-xx-xx" VALUE2="1988-03-26-xx-xx-xx">next
    week</TIMEML>, and some people here are wondering whether its American
    hamburgers will be as popular as the local fast-food treat, Pljeskavica.
</PARAGRAPH>
<PARAGRAPH ID="2" VALUE="0">
```

The long-awaited opening of the restaurant on one of Belgrade's main downtown squares will take place <TIMEML ID="2" TYPE="date" VALUE="1988-03-24-xx-xx-xx">March 24</TIMEML>, the Yugoslav news agency Tanjug reported, and it will offer Big Macs, fries and the other specialties familiar to McDonald's customers in the West.

</PARAGRAPH>

<PARAGRAPH ID="3" VALUE="-1">

The Belgrade media have suggested that the success of the American restaurant depends on its acceptance by Yugoslavians who are long accustomed to the hamburger-like Pljeskavica.

</PARAGRAPH>

<PARAGRAPH ID="4" VALUE="0">

Pljeskavica is made of ground pork and onions, and it is served on bread and eaten with the hands. It is sold at fast-food restaurants across the country and costs about a dollar.

</PARAGRAPH>

<PARAGRAPH ID="5" VALUE="0">

``In fact, this is a clash between the Big Mac and Pljeskavica," said an official of Genex, Yugoslavia's largest state-run enterprise that will operate the McDonald's.

</PARAGRAPH>

<PARAGRAPH ID="6" VALUE="0">

John Onoda, a spokesman at McDonald's Oak Brook, Ill., headquarters, said it was the first of the chain's outlets in a communist country.

</PARAGRAPH>

<PARAGRAPH ID="7" VALUE="+1">

The next East European McDonald's is scheduled to be opened in Budapest, Hungary, <TIMEML ID="3" TYPE="date" VALUE="1988-03-24-xx-xx-xx"

```
        VALUE2="1988-12-31-xx-xx-xx">by the end of this year</TIMEML>, said Vesna
        Milosevic, another Genex official.
    </PARAGRAPH>
    <PARAGRAPH ID="8" VALUE="-2">
        Negotiations have been going on for years for expanding the fast-food chain to
        the Soviet Union, but no agreement has been announced.
    </PARAGRAPH>
```

3.2.3.7 Time Structure Chart

After we build up the time structure of the news story, we get values for all the paragraphs and then convert to a time structure chart for the news story. To find out any possible patterns, we need to compare all the time structures of these news stories. It is much easier to compare shapes than massive numbers. In a time structure chart, axis x means paragraph and axis y means story time of each paragraph. Figure 3.1 is an example chart representing time structure of McDonald's Opening story.

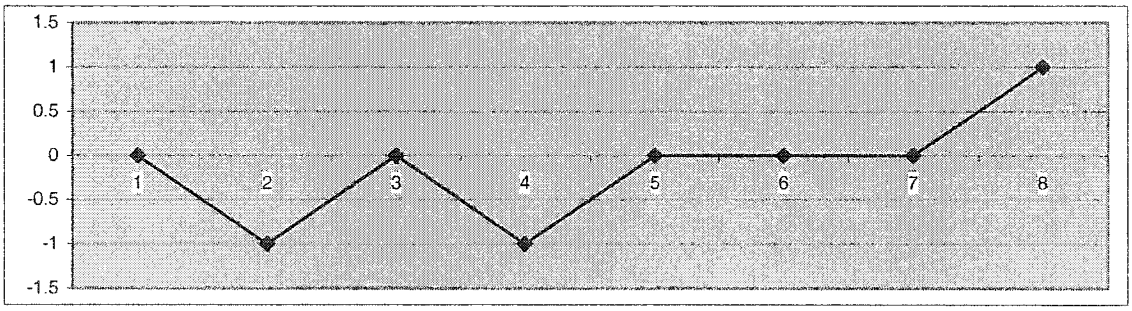

Figure 3.1: Time structure of McDonald's opening story

The chart above tells us that it is an eight-paragraph story. There are five paragraphs in this story having the story time zero in which the main event happened. Two paragraphs have the story time before main event (with negative values), probably introducing the background about the main event. One paragraph has the story time after main event (with positive value), probably providing follow-up or people's reactions.

3.3 Analysis of Results

In this project we have studied 62 news stories. The stories cover various topics including politics, economics, disasters, science, entertainment and terror attacks. After analyzing all the target texts, we annotated the texts and derived their corresponding time structures. We now analyze these time structures to see if any interesting patterns may exist. The annotated texts are useful in future projects.

3.3.1 Analyzing the Time Structure

When analyzing the time structure, we ignore the paragraphs with Time zero, only count the number of paragraphs with the story time before main event (above the axis x in chart) and the story time after main event (below the axis x in chart). We categorize time structures into different groups according to the shape of structures. The groups are defined as below:

- **All Before (AB)**: A time structure has all paragraphs with negative values or Time zero.
- **Most Before (MB):** A time structure has all paragraphs with negative value or Time zero except one paragraph.

- **Mixed (M):** A time structure has paragraphs with mixed value of negative number, Time zero and positive number.
- **Most After (MA):** A time structure has all paragraphs with positive value or Time zero except one paragraph.
- **All After (AA):** A time structure has all paragraphs with positive values or Time zero.

The overall results are listed in Table 3.4.

Time structure group	Number of texts	Percentage
AB	23	37.1%
MB	14	22.6%
M	22	35.5%
MA	1	1.6%
AA	2	3.2%

Table 3.4: Average story time of news stories

Among all the texts, 37.1% (23 of 62) describe only events that happened before main event or roughly at the same time. 22.6% of texts (14 of 62) that describe only events that happened before main event or roughly at the same time, except one event that happened after main event, or even not happen yet. A total of 59.7% news stories mostly tell readers about background of the main event, cause of main event, even recall some history if there are similar events happened in the past.

Only 4.8% of the texts (3 of 62) describe events mostly after main event or roughly at the same time as main event. These news stories only focus on people's reaction on main event, impact of main event or plans in future.

35.5% of the texts (22 of 62) have mixed events before and after main event of the story. As we see in the example text, the story describes negotiations in the past. It also reveals the plan of opening the next East European McDonald's in the future.

The news stories we analyzed cover various topics such as politics, economics, sciences etc. If we take a close look into those different categories, we can find that the results are slightly different among categories. The results are listed in Table 3.5.

Category	**Politic**		**Economic**		**Science**		**Terror Attack**		**Disaster**		**Other**	
# of texts	26		9		6		6		7		8	
	#[1]	%[2]	#	%	#	%	#	%	#	%	#	%
AB	9	34.6	2	22.2	3	50	1	16.7	2	28.6	6	75
MB	9	34.6	2	22.2	0	0	1	16.7	1	14.3	1	12.5
M	7	26.9	4	44.5	3	50	3	50	4	57.1	1	12.5
MA	0	0	0	0	0	0	1	16.6	0	0	0	0
AA	1	3.9	1	11.1	0	0	0	0	0	0	0	0

Table 3.5: Average story time in each pattern group

Note 1: Number of texts in each of the pattern group

Note 2: Percentage of each pattern group

The statistics shows that the percentages in each pattern group are different. The percentage of texts that belong to AB (All Before), MB (Most Before) and M (Mixed) are greater than any other pattern groups in each of the topic categories.

3.3.2 Discussion

According to data we derived from section 3.3.1, we can see that a total of 95.2% texts (59 of 62) fall into three pattern groups, AB, MB, M. The percentages for each pattern group are 37.1%, 22.6%, 35.5%, respectively. This result is very indicative. As we pre-set 20% as a threshold for being recognized as a pattern, we can conclude that there exist patterns that journalists used to organize events in the news stories. AB. MB and M are considered as patterns in our study. These three patterns are held in all the categories when we take a detailed look into each of the topic categories.

News stories are a rich source of language that reflects our everyday life. The study of time structure patterns of news stories enables us to better understand how journalists organize our daily happenings into news stories. Thus it helps us to understand the functioning of language of society. The study also enables us to compare news stories with other kinds of stories, such as personal narratives, in terms of time structure pattern. We can also compare news stories with other media genres such as newscasts on radios or televisions.

Paragraphs are used for separating ideas that authors intend to describe so that things or events in the same paragraph usually are closely related. With this function of paragraph, this study also possibly enables us to extract implicit temporal information. The time of an event might be determined by the implication from the event in the same paragraph. To confirm this, more work has to be done in future. The result might benefit applications in IE area.

CHAPTER 4

CONCLUSION AND FUTURE WORK

News articles are a particular type of story that reflects our daily life on various media. The most distinctive characteristics of news stories are its discontinuity in the ordering of events and its limited shelf life. These characteristics present difficulties for automated systems which deal with temporal information. Such applications include information extraction, question answering, summarization and machine translation, etc.

Corpus-driven approaches provide frameworks and mechanisms to those systems to extract temporal information from text. The demands of annotated corpus for the approaches promote the development of thorough yet effective annotation schemes. TimeML is a XML mark-up language that offers the most thorough scheme so far for identifying and annotating events, times and their relations in text. The annotation scheme is greatly beneficial to those information extraction applications. However the issues about implicit temporal information and temporal relations still remain unsolved.

This major report presents an annotation scheme used to annotate paragraphs and temporal expressions in news texts. The annotation scheme is evaluated by applying to a corpus used for DUC-2002. The whole study took place in two stages: constructing time structures and analyzing time structures. The first stage consists of several steps:

annotating all temporal expressions reported in text, annotating all paragraphs, and building time structure based on paragraphs. These steps were repeated in every news text being analyzed until we got all time structures in the corpus we have chosen. This stage of work also resulted in an annotated corpus that will be used in future study and this will be discussed in section 4.1.

The second stage is to analyze time structures derived from the first stage. The analysis also involves several steps: comparing all 62 time structure charts generated, classifying them into different groups, and calculating the percentage of each group against the total number of analyzed news texts.

To be recognized as a pattern, a kind of time structure must appear over a pre-set threshold, say 20% of all structures. The analysis result from the second stage shows that there exist patterns that journalists used to construct news stories. There are 37.1% news stories that describe events which all happened before the main event or roughly at the same time as the main event. 22.6% of news stories describe events that mostly happened before the main event or roughly at the same time as the main event. 35.5% of news stories have mixed events where some happened before the main event and some happened or will happen after the main event. These three types of time structure are considered as patterns, and called as All Before (AB), Most Before (MB) and Mixed (M) respectively. There are only 4.8% of news stories that describe follow-up events to the main event or events that will happen in the near future. This phenomenon reflects the fact that news agencies pursue the news value of recency.

The stories we analyzed cover various topics such as politics, economics, sciences, disasters etc. Although the percentages of texts in pattern groups vary among different categories, patterns are still recognizable in each of the topic categories. AB, MB and M are the patterns held in all the topic categories we have studied.

Paragraphs serve the function of separators for ideas that authors intend to describe. Things or events described in the same paragraph usually are closely related. Sometimes events in the same paragraph happened in sequence. This brings us an idea for a possible way to capture temporal information that can not be captured using currently existing temporal annotation schemes.

It is difficult to determine relations between two events without having a temporal relation indicator between them. But in a news story, there are often situations that two events described in a paragraph have no explicit temporal relation. If we can't locate times for both events, is it possible that we can determine the time for one event according to another event? Can we say that two events in the same paragraph have same time?

The following example is a news story titled "Honecker Arrested, Taken to Prison" from the American Press, March 14, 1998 edition.

> Honecker was immediately taken to East Berlin's Rummelsburg prison, the brief dispatch said. Honecker joins other members of his ousted politburo already in prison awaiting trial.

The time for the first event "taken" can be located by the context using temporal expression "immediately", but the time for second event "joins" is no way to determine for an automated system without a sophisticated semantic lexicon. This is a case of above situation that no explicit relation exists between two events. For humans it is easy to understand that the second event is an immediate result caused by the first event so that it is safe to assign to the second event the same time value as the first event. Can we say this is a phenomenon? How reliable can this method be? An answer can be given only after a more detailed investigation.

All the annotation of temporal expressions and paragraphs are done manually in this study. This is not practical to apply the annotation scheme to a large scale corpus. In the future some existing systems will be used for temporal expression identification.

Other possible related work in the future are to apply the scheme to texts in other domains or genres. Domain like financial news might have a very different variety of mechanisms used to convey temporal information. Different patterns of time structure might be found in composition of news articles.

REFERENCES

Allen, J. (1983), "Maintaining Knowledge about Temporal Intervals", *Communications of the ACM,* 26:832-843.

Allen, J. (1991), "Time and Time Again: The Many Ways to Represent Time", *International Journal of Intelligent Systems* 6(4).

Bell, A. (1995a), "News Time", *Time & Society,* 4, 305-28.

Bell, A. (1995b), "Language and the Media", *Annual Review of Applied Linguistics,* 15, 23-41.

Bell, A.; Garrett, P. (1997) "The Discourse Structure of News Stories", *Approaches to Media Discourse,* (Chapter 3), Published by Blackwell Publishers.

Dean, T.; McDermott, D. (1987) "Temporal Data Base Management", *Journal of Artificial Intelligence* (32), pp 1-55.

Ferro, L.; Mani, I.; Sundheim, B.; Wilson, G. (2001) "TIDES Temporal Annotation Guidelines", *MITRE technical Report,* MTR 01W0000041.

Hobbs, J.; Pustejovsky, J. (2003) "Annotating and Reasoning about Time and Events", *Working Papers of the 2003 AAAI Spring Symposium on Logical Formalization of Commonsense Reasoning*, pp 74-82.

Labov W.; Waletzky J. (1967) "Narrative Analysis: Oral Versions of Personal Experience", *Essays on the Verbal and Visual Arts (Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society),* pp. 12-24.

Mehrotra, K.; Chai, J.; Pillutla, S (1991) "A Study of Approximating the Moments of the Job in PERT Networks", *Technical Report, School of Computer and Information Science, Syracuse University.*

Ohtsuka, K.; Brewer, W.F.; (1992) "Discourse Organization in the Comprehensin of Temporal Order in Narrative Texts", *Discourse Processes* 15(3).

Pustejovsky J.; Saurí R.; Setzer A.; Gaizauskas B.; Ingria B. (2002), "TimeML Annotation Guidelines", *TERQAS Workshop document.*

Setzer, A. (2001), "Temporal Information in Newswire Articles: An Annotation Scheme and Corpus Study", *PhD Thesis, University of Sheffield.*

Vilain, M.B. (1982), "A System for Reasoning about Time", *Proceedings of AAAI-82,* pp. 197-201.

APPENDIX A: ANNOTATED TEXT EXAMPLES

Some examples of annotated news texts:

Example Text 1:
First McDonald's to Open in Communist Country (d064j\AP880314-0110)

```
<DOC>
<DOCNO> AP880314-0110 </DOCNO>
<FILEID>AP-NR-03-14-88 1223EST</FILEID>
<FIRST>r i AM-Yugoslavia-McDonalds    03-14 0243</FIRST>
<SECOND>AM-Yugoslavia-McDonalds,0250</SECOND>
<HEAD>First McDonald's to Open in Communist Country</HEAD>
<HEAD>With AM-Gorbachev, Bjt</HEAD>
<DATELINE>BELGRADE, Yugoslavia (AP) </DATELINE>
<TEXT>
  <PARAGRAPH ID="1" VALUE="0, -1">
      The communist world gets its first McDonald's <TIMEML ID="1" TYPE="date" VALUE="1988-03-20-xx-xx-xx" VALUE2="1988-03-26-xx-xx-xx">next week</TIMEML>, and some people here are wondering whether its American hamburgers will be as popular as the local fast-food treat, Pljeskavica.
  </PARAGRAPH>
  <PARAGRAPH ID="2" VALUE="0">
      The long-awaited opening of the restaurant on one of Belgrade's main downtown squares will take place <TIMEML ID="2" TYPE="date" VALUE="1988-03-24-xx-xx-xx">March 24</TIMEML>, the Yugoslav news agency Tanjug reported, and it will offer Big Macs, fries and the other specialities familiar to McDonald's customers in the West.
  </PARAGRAPH>
  <PARAGRAPH ID="3" VALUE="-1">
      The Belgrade media have suggested that the success of the American restaurant depends on its acceptance by Yugoslavians who are long accustomed to the hamburger-like Pljeskavica.
  </PARAGRAPH>
  <PARAGRAPH ID="4" VALUE="0">
      Pljeskavica is made of ground pork and onions, and it is served on bread and eaten with the hands. It is sold at fast-food restaurants across the country and costs about a dollar.
  </PARAGRAPH>
  <PARAGRAPH ID="5" VALUE="0">
      ``In fact, this is a clash between the Big Mac and Pljeskavica," said an official of Genex, Yugoslavia's largest state-run enterprise that will operate the McDonald's.
  </PARAGRAPH>
  <PARAGRAPH ID="6" VALUE="0">
```

```
      John Onoda, a spokesman at McDonald's Oak Brook, Ill., headquarters, said it
      was the first of the chain's outlets in a communist country.
  </PARAGRAPH>
  <PARAGRAPH ID="7" VALUE="+1">
      The next East European McDonald's is scheduled to be opened in Budapest,
      Hungary, <TIMEML ID="3" TYPE="date" VALUE="1988-03-14-xx-xx-xx"
      VALUE2="1988-12-31-xx-xx-xx">by the end of this year</TIMEML>, said
      Vesna Milosevic, another Genex official.
  </PARAGRAPH>
  <PARAGRAPH ID="8" VALUE="-2">
      Negotiations have been going on for years for expanding the fast-food chain to
      the Soviet Union, but no agreement has been announced.
  </PARAGRAPH>
</TEXT>
</DOC>
```

Example Text 2.
Report: Honecker Unlikely to Go to Trial in East Germany (d07f\AP900825-0099)

```
<DOC>
<DOCNO> AP900825-0099 </DOCNO>
<FILEID>AP-NR-08-25-90 1720EST</FILEID>
<FIRST>r i AM-EastGermany-Honecker     08-25 0263</FIRST>
<SECOND>AM-East Germany-Honecker,0270</SECOND>
<HEAD>Report: Honecker Unlikely To Go to Trial in East Germany</HEAD>
<DATELINE>WEST BERLIN (AP) </DATELINE>
<TEXT>
  <PARAGRAPH ID="1" VALUE="0">
      Ousted East German leader Erich Honecker will not stand trial in East Germany
      as long as the formerly Communist country exists, a West German newspaper
      reported.
  </PARAGRAPH>
  <PARAGRAPH ID="2" VALUE="0,+1">
      The Hamburg-based Bild am Sonntag said <TIMEML ID="1" TYPE="date"
      VALUE="1990-08-25-xx-xx-xx">Saturday</TIMEML> that it would report in
      its <TIMEML ID="2" TYPE="date" VALUE="1990-08-26-xx-xx-xx">
      Sunday</TIMEML> editions that Honecker could be prosecuted in a united
      Germany, however, for violation of property laws.
  </PARAGRAPH>
  <PARAGRAPH ID="3" VALUE="0">
      Bild quoted Guenter Seidel, an East German prosecutor, as saying that Honecker
      had used $42 million for stocking a private housing estate for leaders of the
      former Communist government.
  </PARAGRAPH>
  <PARAGRAPH ID="4" VALUE="0, +2">
```

```
      However, Seidel said that the investigation was not far enough along to determine
      whether charges could be filed against Honecker before East Germany merges
      with West Germany on <TIMEML ID="3" TYPE="date" VALUE="1990-10-03-
      xx-xx-xx">Oct. 3</TIMEML>.
  </PARAGRAPH>
  <PARAGRAPH ID="5" VALUE="0">
      Negotiators are still working out the merger of the two German legal systems.
  </PARAGRAPH>
  <PARAGRAPH ID="6" VALUE="-2,-1,0">
      Honecker, 78, was ousted as East Germany's leader on <TIMEML ID="4"
      TYPE="date" VALUE="1989-10-18-xx-xx-xx">Oct. 18</TIMEML>, paving the
      way for the country's first freely elected government in <TIMEML ID="5"
      TYPE="date" VALUE="1990-03-xx-xx-xx-xx">March</TIMEML>. Honecker
      is in poor health and remains confined to a Soviet military hospital in Beelitz
      outside East Berlin.
  </PARAGRAPH>
  <PARAGRAPH ID="7" VALUE="0">
      He is under investigation on allegations of abuse of power, corruption, harboring
      terrorists and issuing shoot-to-kill orders to prevent East Germans from escaping
      to West Germany when he served as the country's leader.
  </PARAGRAPH>
  <PARAGRAPH ID="8" VALUE="0">
      Bild said that Erich Mielke, the ex-head of East Germany's former secret police,
      was also unlikely to go to court in East Germany.
  </PARAGRAPH>
  <PARAGRAPH ID="9" VALUE="0">
      ``I am at the end. I am a dead man," Bild quoted Mielke, 82, as saying at his last
      interrogation.
  </PARAGRAPH>
</TEXT>
</DOC>
```

Example Text 3:
Jewish Leaders To Meet in Berlin on German Reunification (d069f\AP900321-0057)

```
<DOC>
<DOCNO> AP900321-0057 </DOCNO>
<FILEID>AP-NR-03-21-90 0648EST</FILEID>
<FIRST>r a PM-Jews-Reunification    03-21 0347</FIRST>
<SECOND>PM-Jews-Reunification,0356</SECOND>
<HEAD>Jewish Leaders To Meet in Berlin on German Reunification</HEAD>
<BYLINE>By CATHERINE CROCKER</BYLINE>
<BYLINE>Associated Press Writer</BYLINE>
```

```
<DATELINE>NEW YORK (AP) </DATELINE>
<TEXT>
  <PARAGRAPH ID="1" VALUE="0">
      The Berlin villa where the Nazi leadership met to implement the ``final solution"
      against Europe's Jewish population will be the site of a World Jewish Congress
      conference on German reunification.
  </PARAGRAPH>
  <PARAGRAPH ID="2" VALUE="0,+2">
      The New York-based organization announced <TIMEML ID="1" TYPE="date"
      VALUE="1990-03-21-xx-xx-xx">Tuesday</TIMEML> that the meeting is
      scheduled for <TIMEML ID="2" TYPE="date" VALUE="1990-05-08-xx-xx-
      xx">May 8</TIMEML>, the 45th anniversary of V-E Day.
  </PARAGRAPH>
  <PARAGRAPH ID="3" VALUE="0">
      ``The world is expecting a Jewish statement on reunification," said Edgar
      Bronfman, the president of the World Jewish Congress.
  </PARAGRAPH>
  <PARAGRAPH ID="4" VALUE="0">``I think that the place where they decided on
      the `final solution' is the right place to have such a meeting, and the anniversary
      of the victory in Europe seems to be the right day."
  </PARAGRAPH>
  <PARAGRAPH ID="5" VALUE="+2,-1">
      Some 100 members of the World Jewish Congress from about 70 countries will
      attend the conference at the villa in Wannsee, Berlin, where on <TIMEML
      ID="3"      TYPE="date"      VALUE="1942-01-20-xx-xx-xx">Jan.      20,
      1942</TIMEML>, high German government and party officials adopted a plan to
      exterminate the Jews.
  </PARAGRAPH>
  <PARAGRAPH ID="6" VALUE="0">
      ``Jews have a right to state a moral position on German reunification and we will
      make a moral statement. And part of that moral statement is to see that it (the
      Holocaust) doesn't happen again," Bronfman said.
  </PARAGRAPH>
  <PARAGRAPH ID="7" VALUE="0">
      In the long run, he said, a united Germany can be ``a very positive thing" for
      Europe, but certain precautions must be taken to prevent a repeat of history. For
      example, Germany's borders ``must be sealed in cement," and the country should
      not be allowed nuclear weapons, he said.
  </PARAGRAPH>
  <PARAGRAPH ID="8" VALUE="0">
      Bronfman said he does not fear a resurgence of anti-Semitism once the two
      Germanys are united. In fact, he said, reunification might cause a decline in anti-
      Jewish feelings.
  </PARAGRAPH>
  <PARAGRAPH ID="9" VALUE="0">
```

```
      B``Now that reunification is about to be a fact, they won't be blaming the Jews" for a divided Germany, Bronfman explained.
  </PARAGRAPH>
  <PARAGRAPH ID="10" VALUE="+1,+2,-2">
      BThe conference will convene <TIMEML ID="3" TYPE="date" VALUE="1990-05-06-xx-xx-xx">May 6</TIMEML> and a formal declaration on German reunification will be issued <TIMEML ID="5" TYPE="date" VALUE="1990-05-08-xx-xx-xx">May 8</TIMEML>. It will be the first meeting ever held on German soil by the World Jewish Congress, which was founded in <TIMEML ID="6" TYPE="date" VALUE="1936-xx-xx-xx-xx-xx">1936</TIMEML> in response to Hitler's threats against Jews.
  </PARAGRAPH>
</TEXT>
</DOC>
```

APPENDIX B: TIME STRUCTURE CHARTS

Charts of temporal structures of the news stories that we have studied:

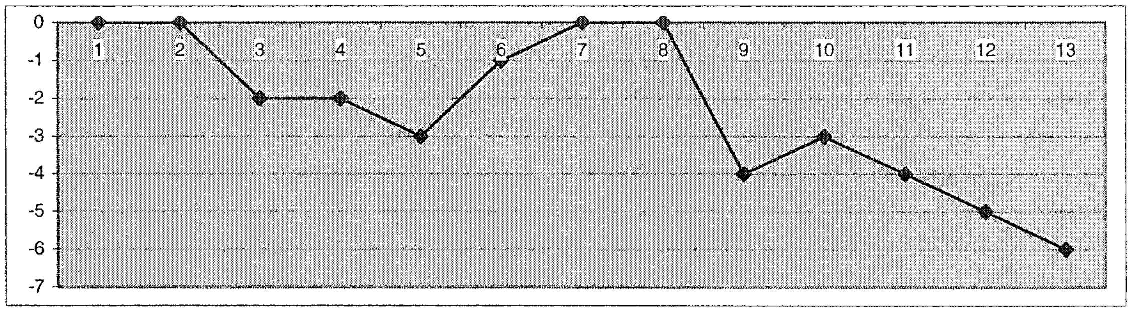

Figure A1: d061j\AP880911-0016

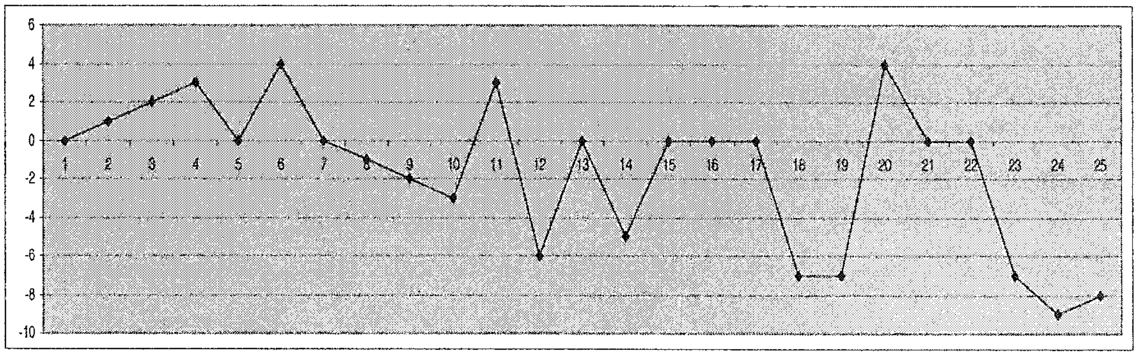

Figure A2: d061j\AP880912-0095

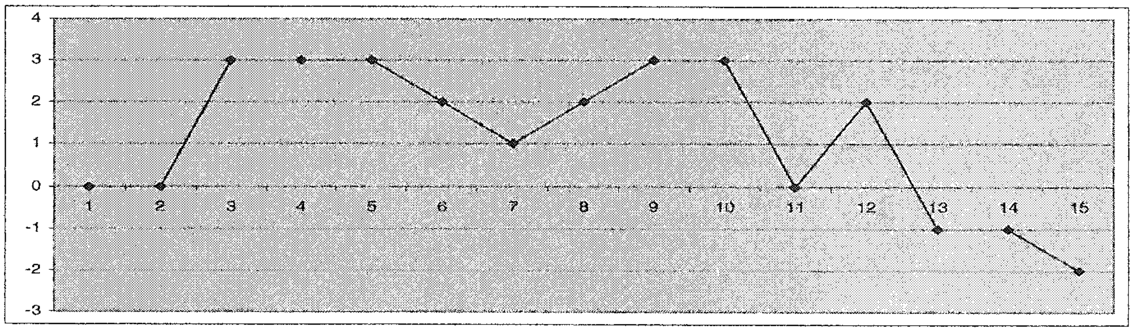

Figure A3: d063j\AP890922-0071

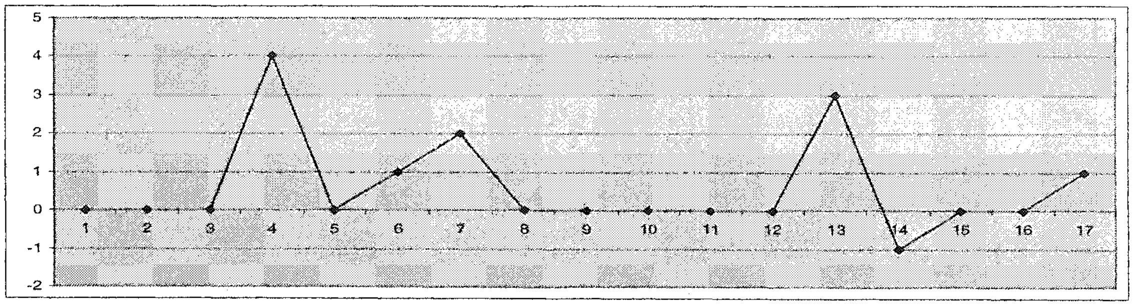

Figure A4: d063j\AP890922-0103

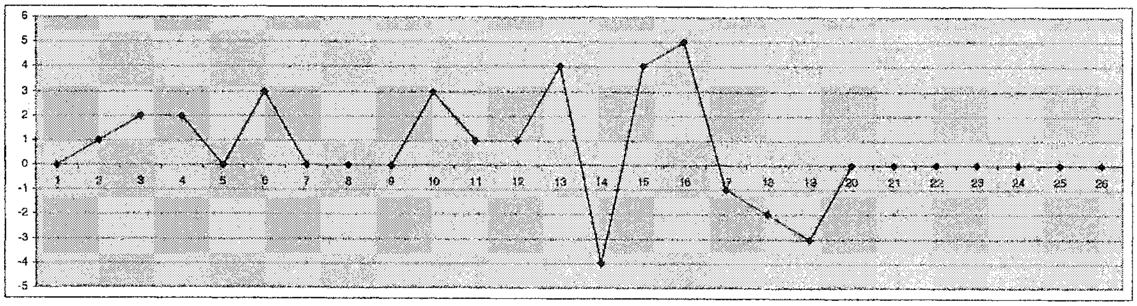

Figure A5: d063j\AP890922-0117

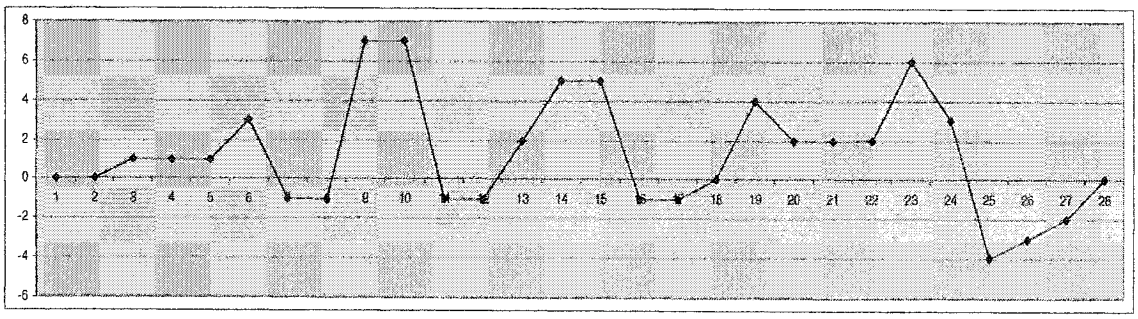

Figure A6: d063j\AP890923-0012

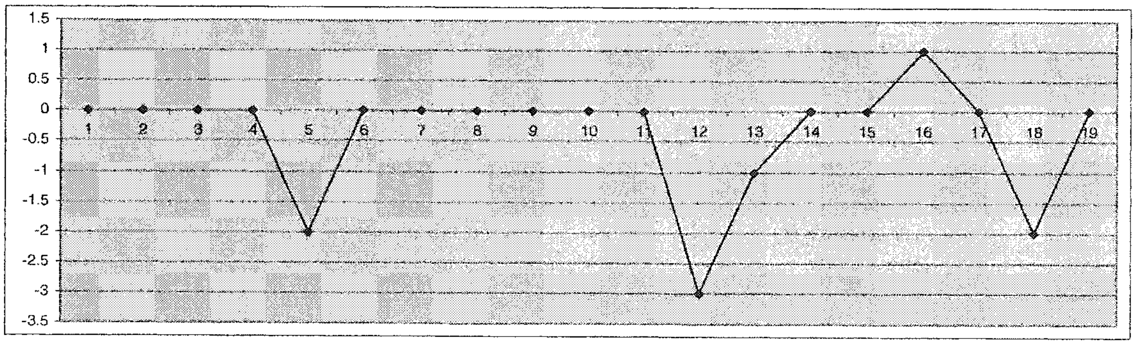

Figure A7: d063j\AP890923-0091

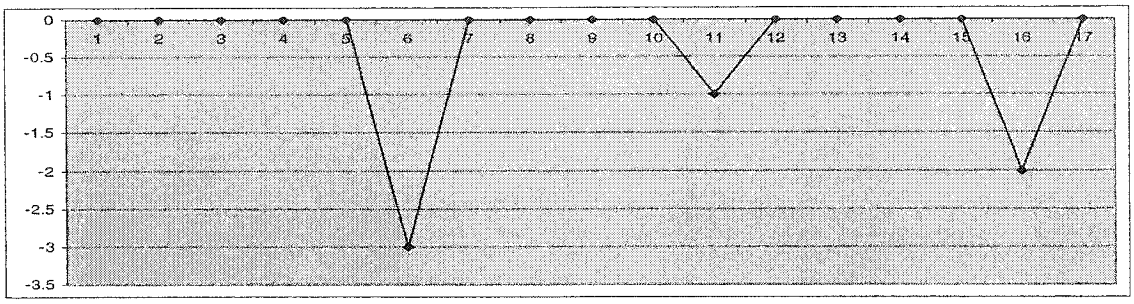

Figure A8: d063j\AP890924-0025

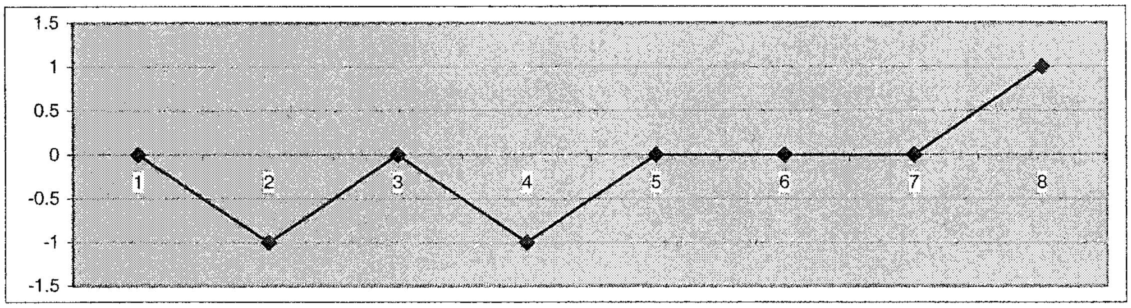

Figure A9: d064j\880314-0110

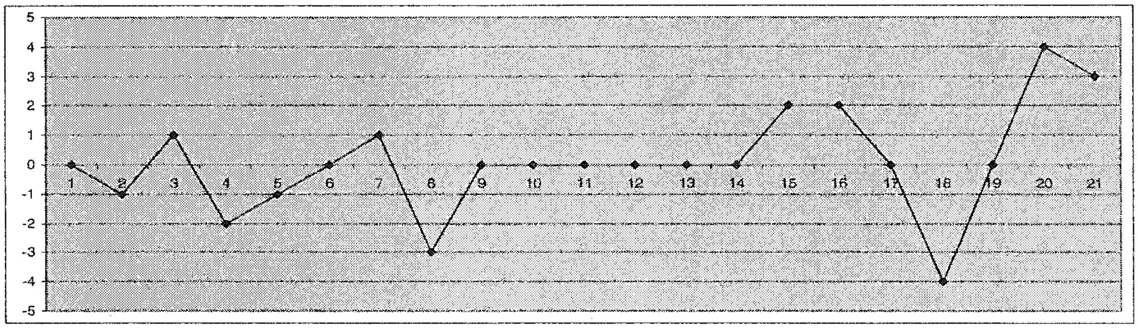

Figure A10: d064j\880324-0193

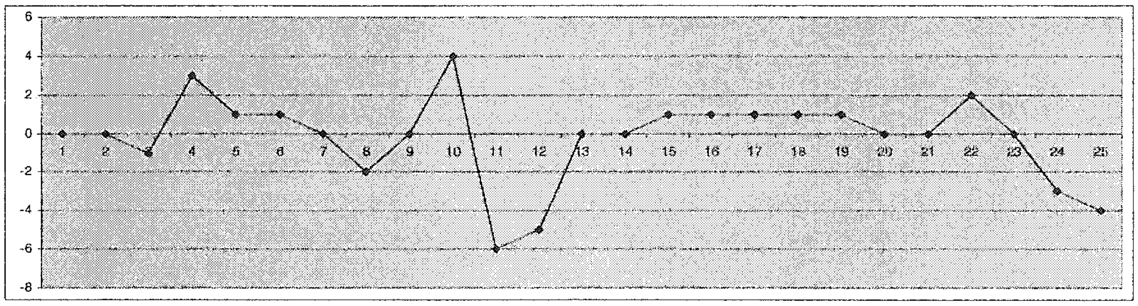

Figure A11: d064j\900131-0200

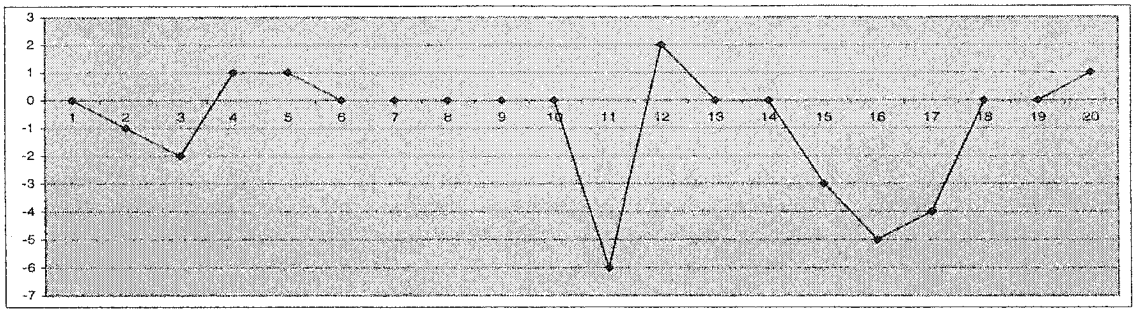

Figure A12: d064j\901008-0136

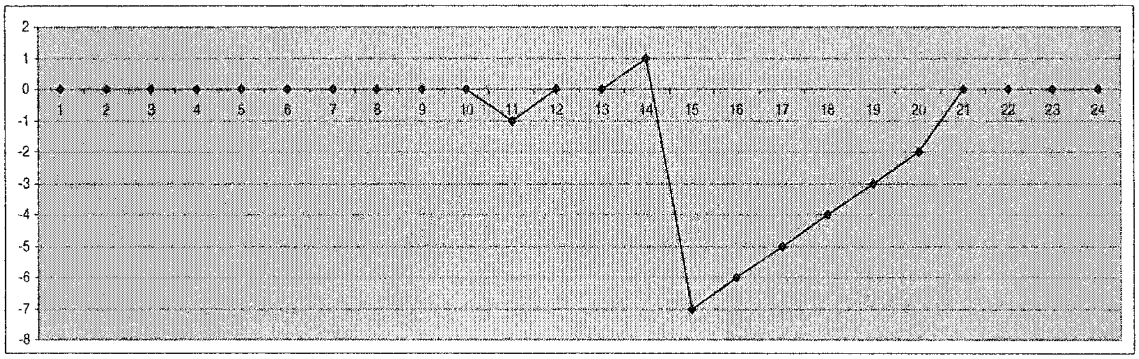

Figure A13: d066j\AP890323-0234

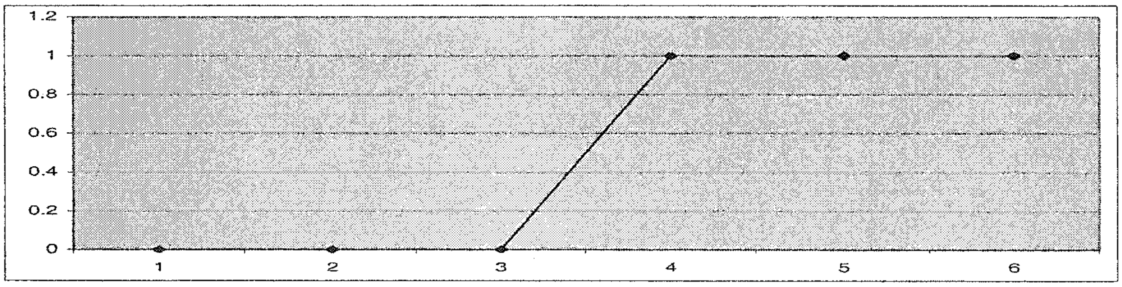

Figure A14: d066j\AP890615-0263

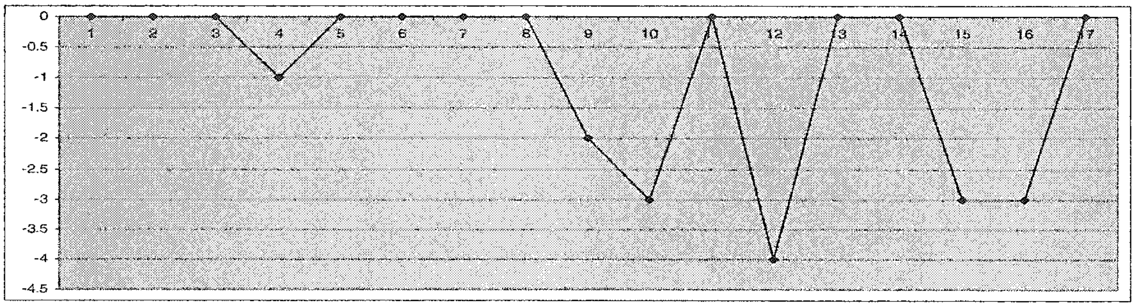

Figure A15: d066j\AP890711-0223

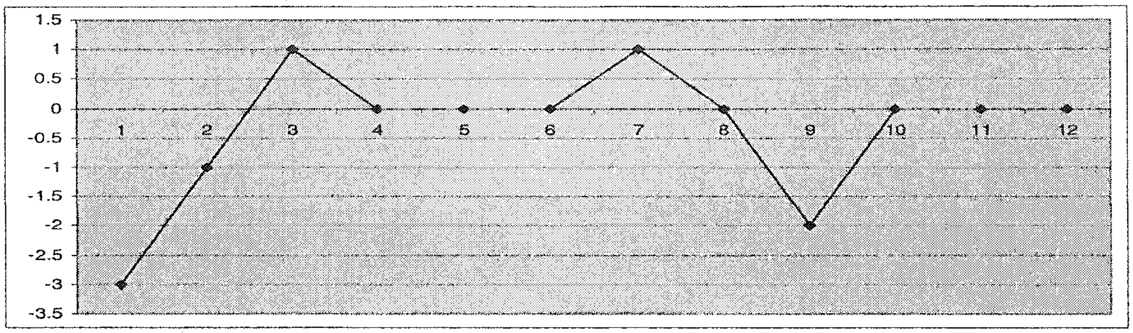

Figure A16: d066j\FT922-14265

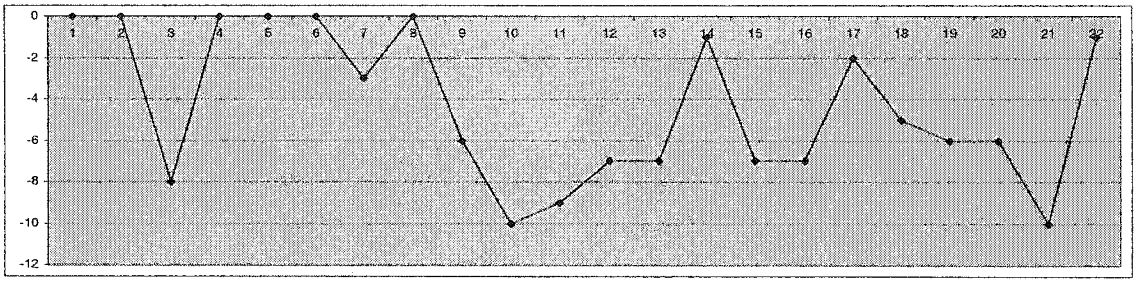

Figure A17: d066j\AP900507-0207

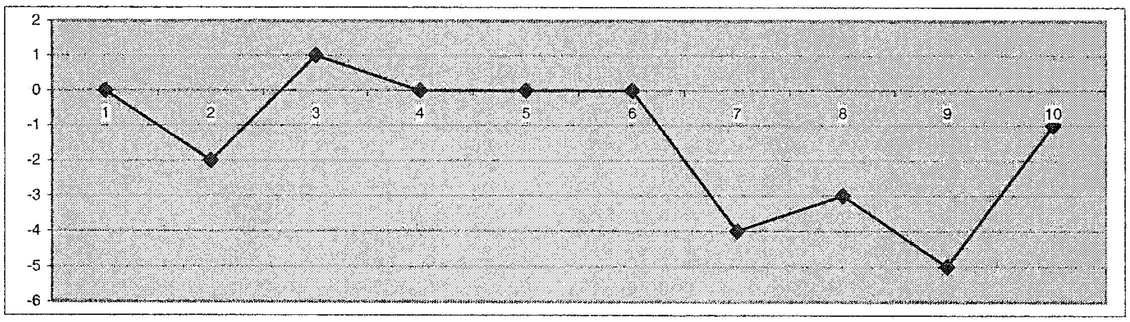

Figure A18: d068f\AP900621-0192

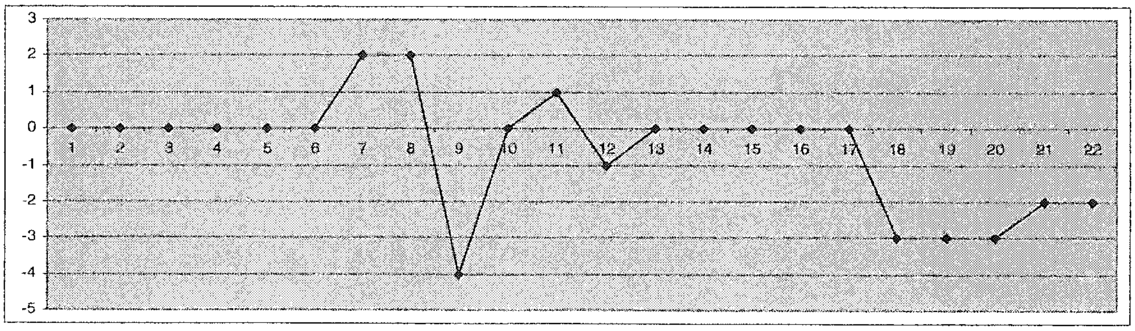

Figure A19: d068f\AP900622-0080

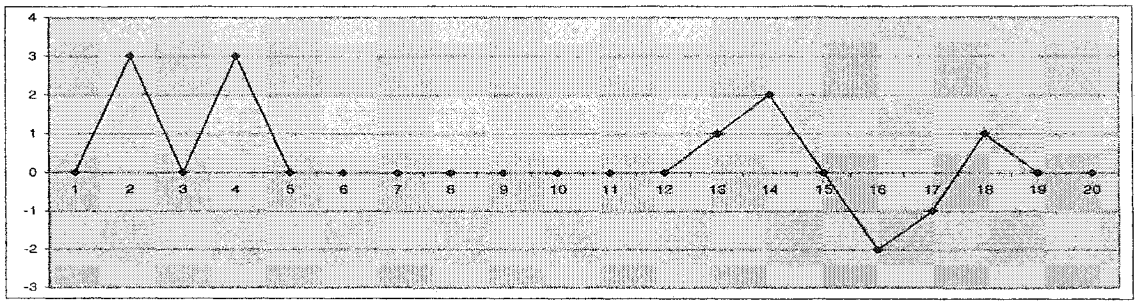

Figure A20: d068f\AP900622-0087

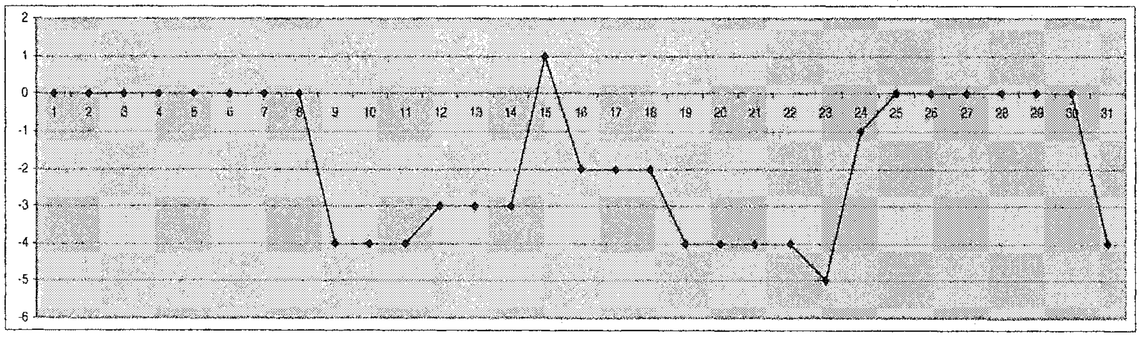

Figure A21: d068f\AP900623-0009

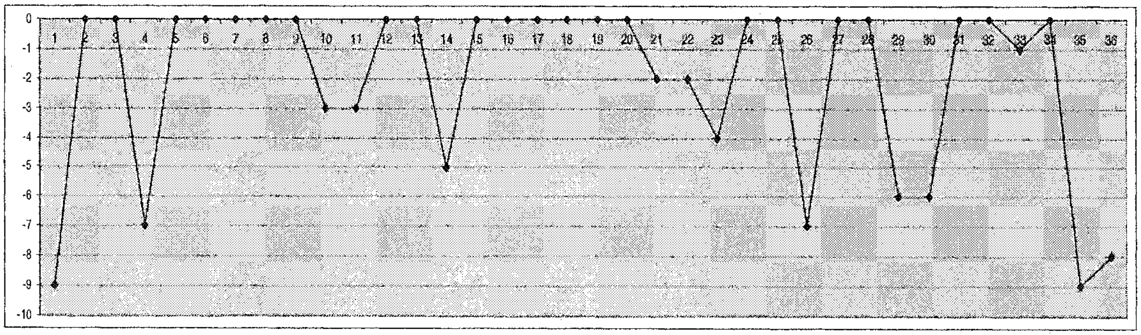

Figure A22: d069f\AP890925-0009

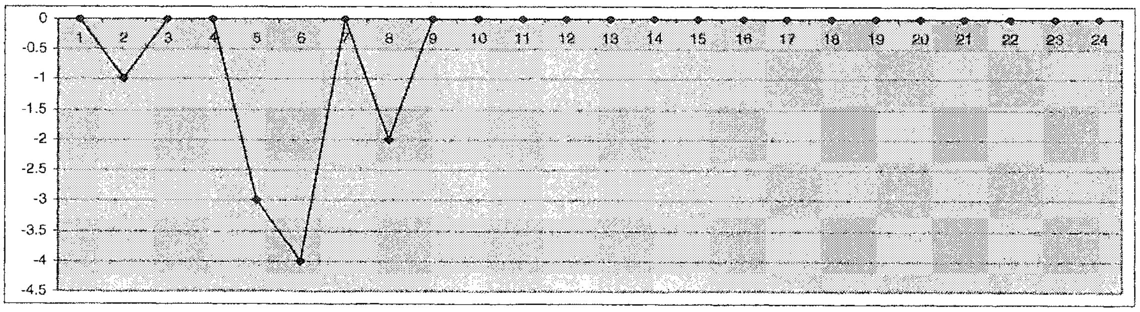

Figure A23: d069f\AP891111-0064

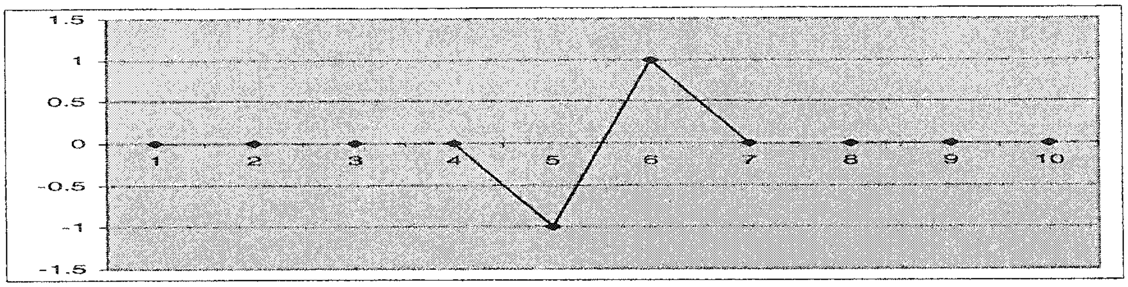

Figure A24: d069f\AP891202-0154

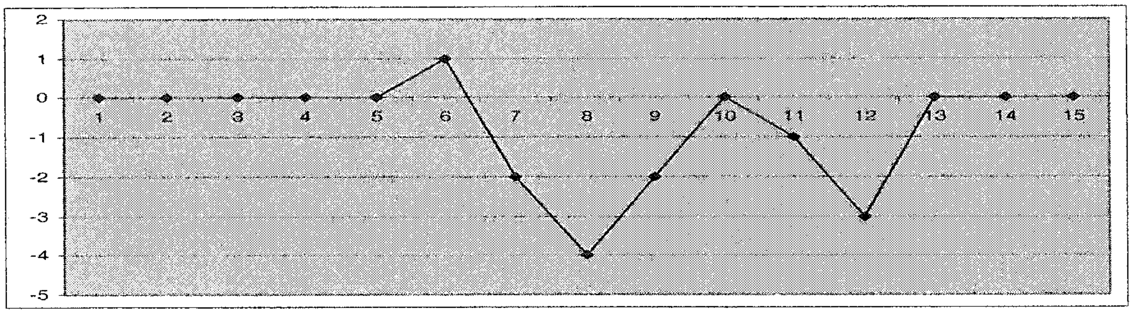

Figure A25: d069f\AP891207-0158

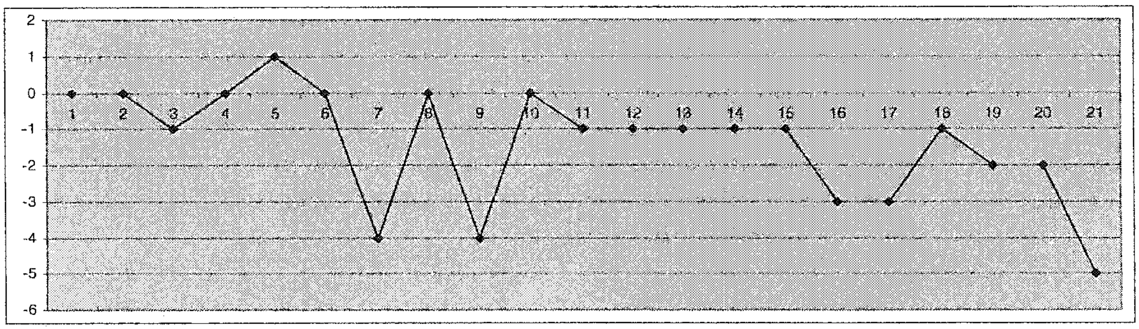

Figure A26: d069f\AP891212-0062

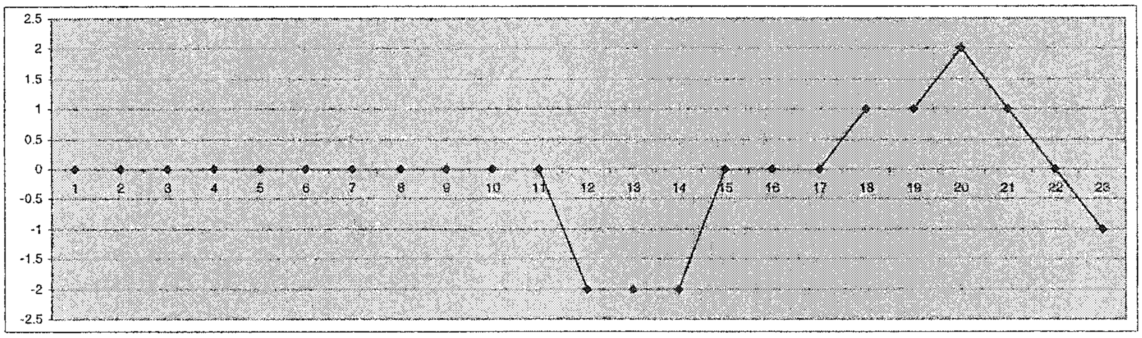

Figure A27: d069f\AP90-030-0202

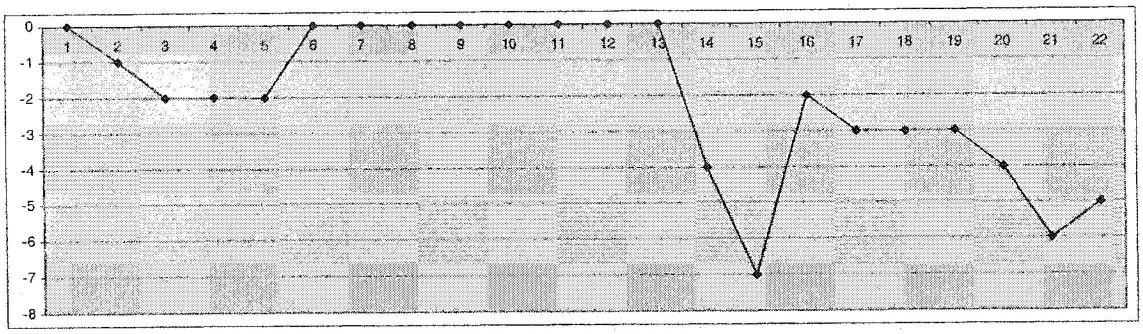

Figure A28: d069f\AP900131-0068

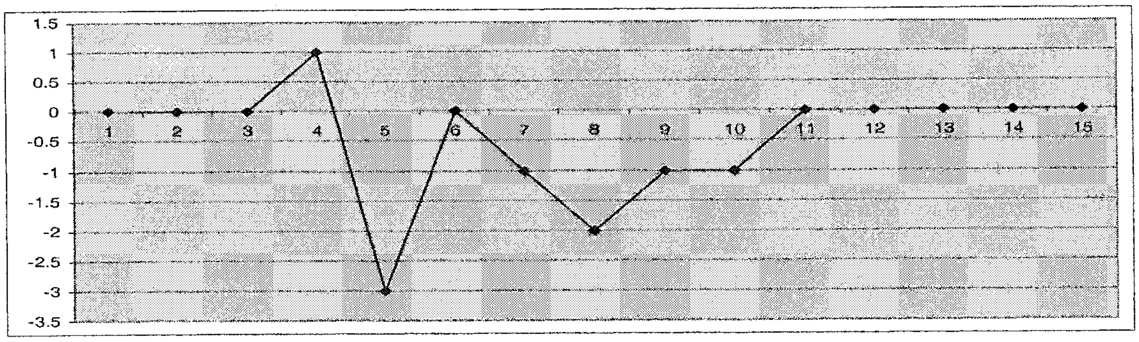

Figure A29: d069f\AP900210-0106

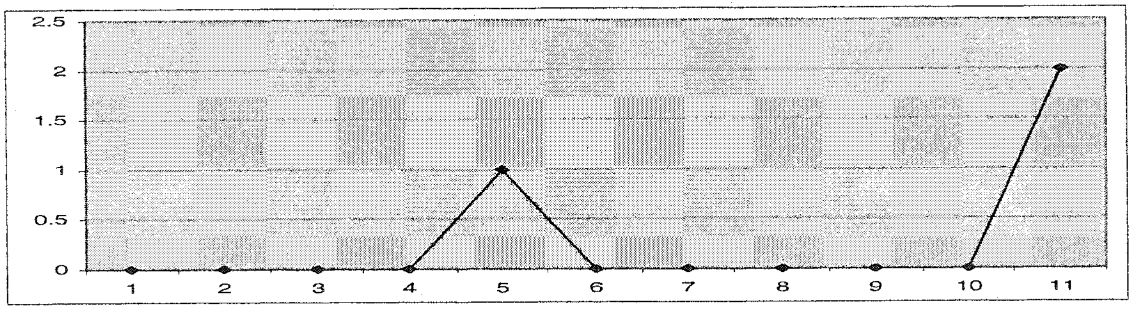

Figure A30: d069f\AP900214-0157

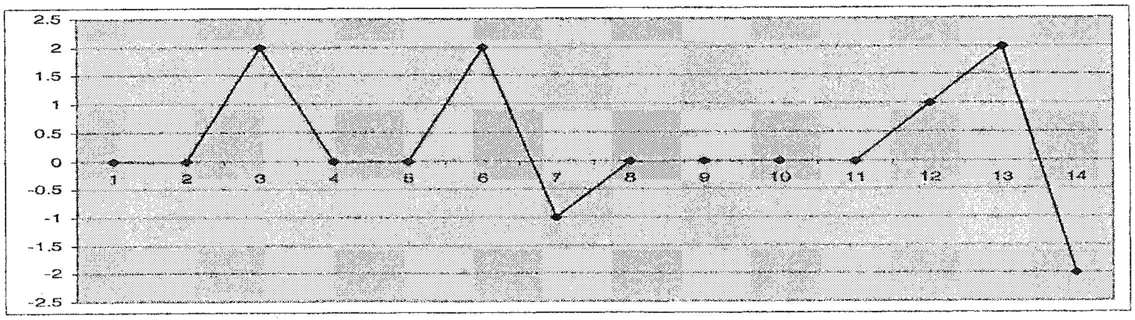

Figure A31: d069f\AP900321-0057

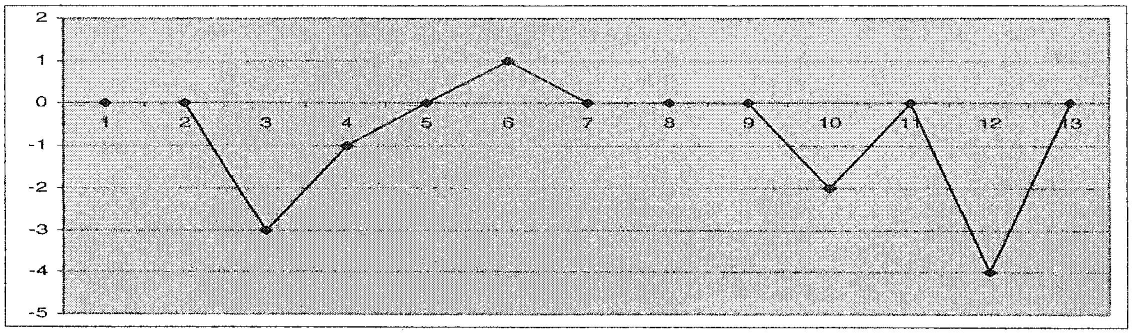

Figure A32: d070f\AP900103-0077

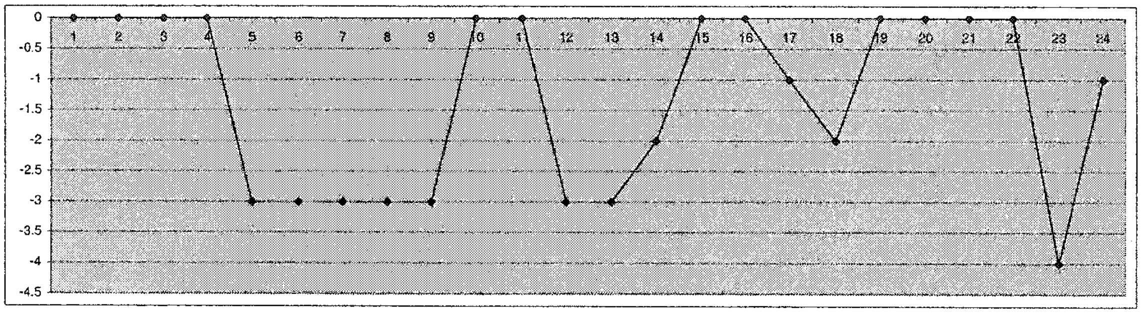

Figure A33: d070f\AP900118-0029

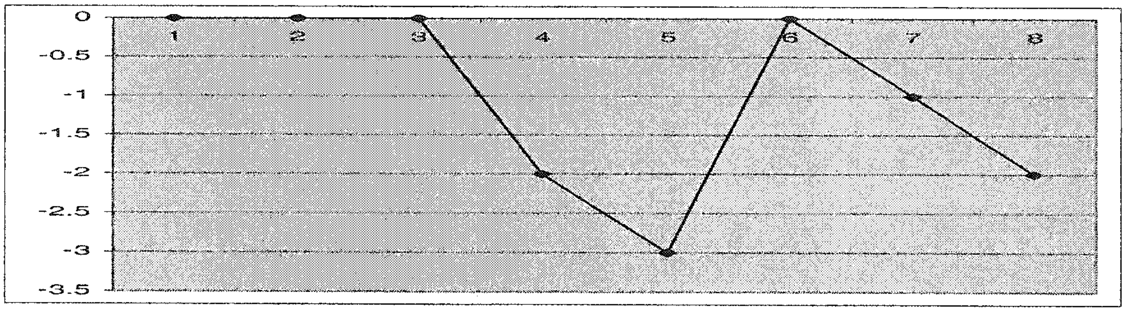

Figure A34: d070f\AP900129-0036

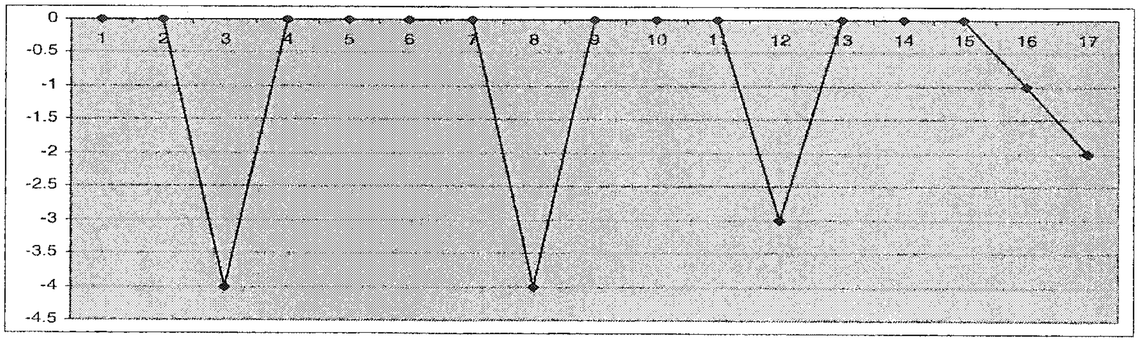

Figure A35: d070f\AP900129-0071

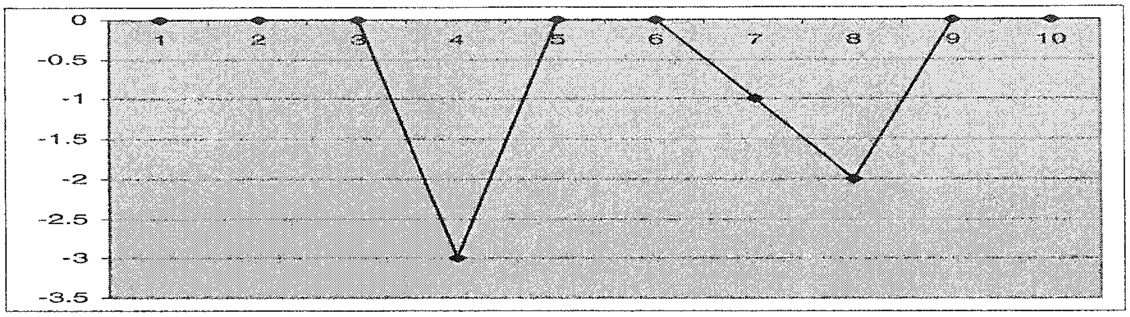

Figure A36: d070f\AP900730-0116

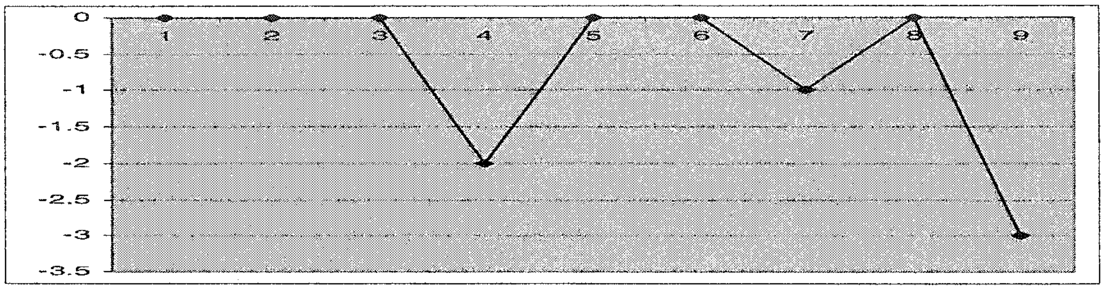

Figure A37: d070f\AP900810-0120

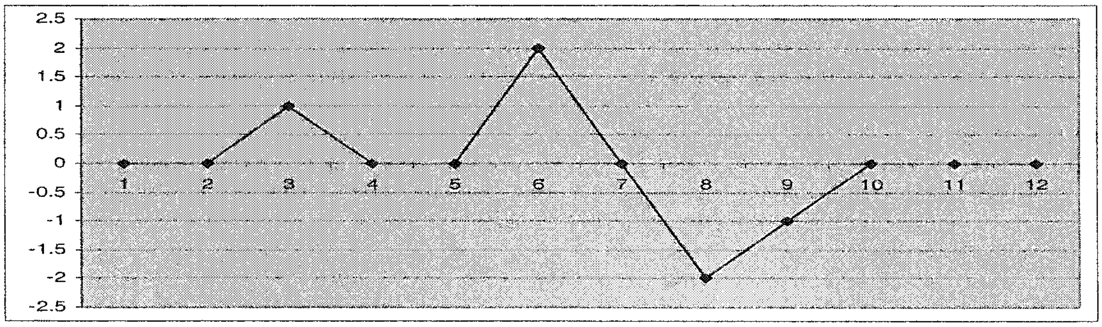

Figure A38: d070f\AP9008125-0099

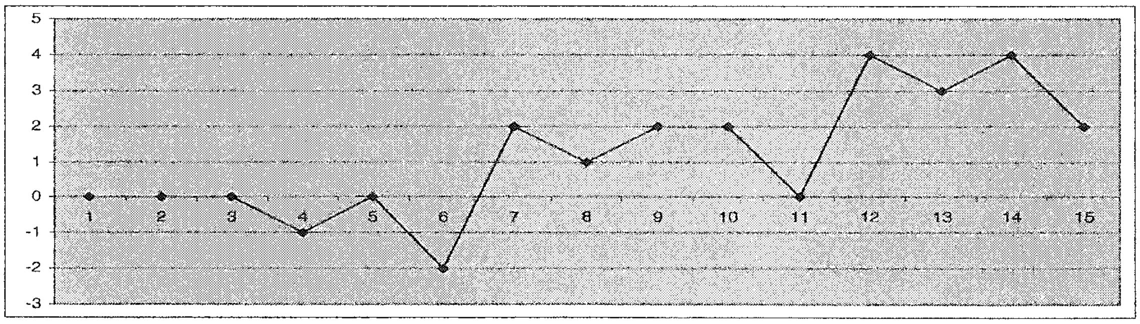

Figure A39: d070f\AP900919-0122

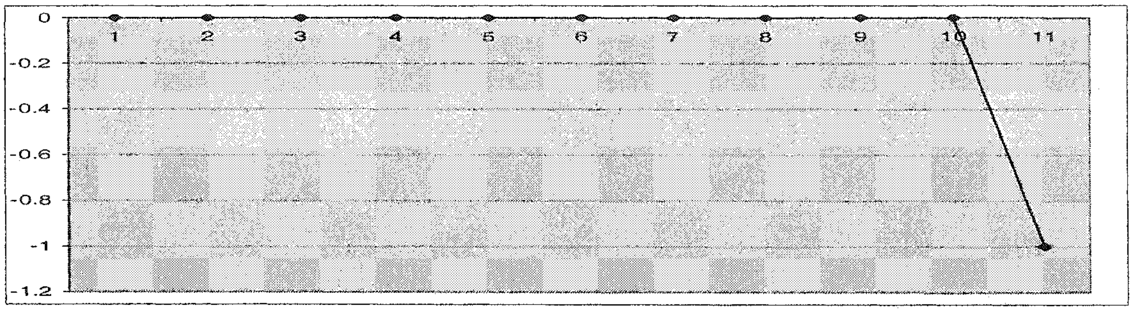

Figure A40: d077b\AP891017-0195

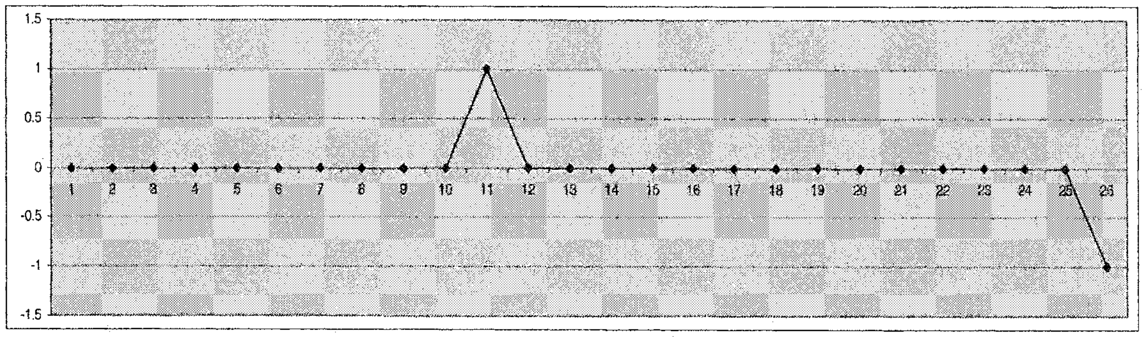

Figure A41: d077b\AP891017-0199

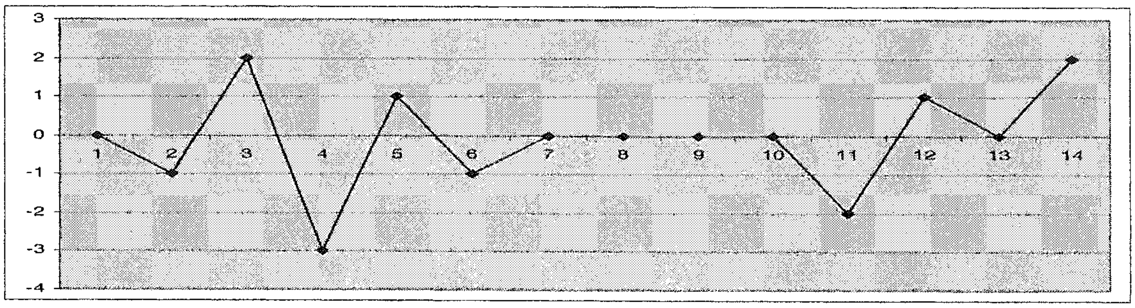

Figure A42: d077b\AP891017-0204

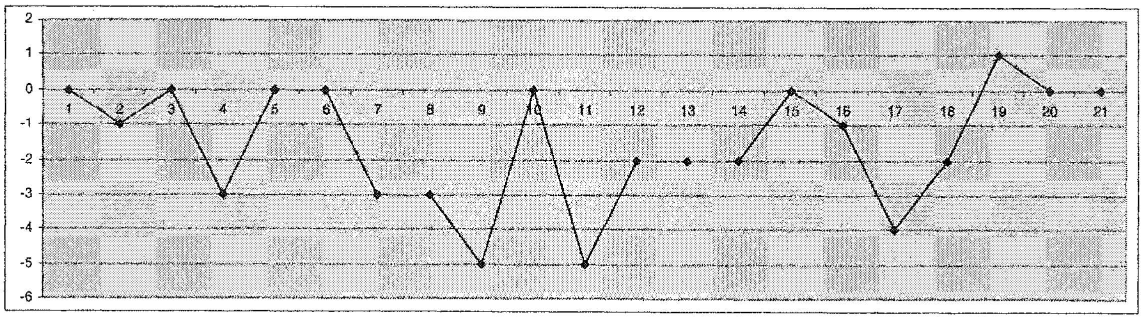

Figure A43: d077b\AP891018-0084

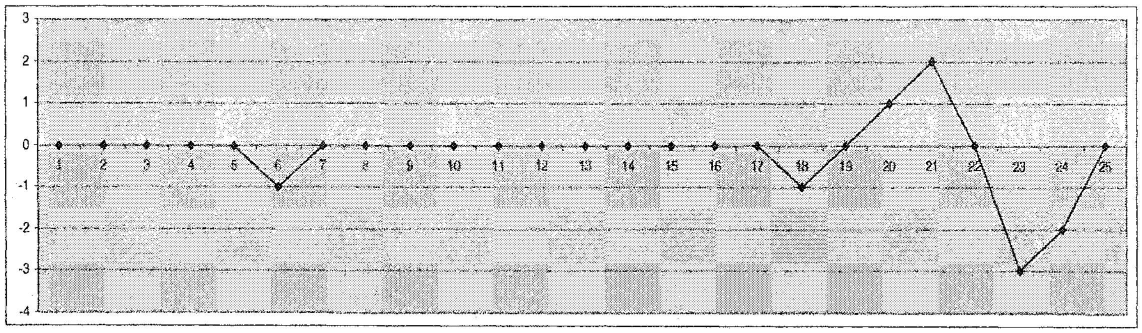

Figure A44: d077b\AP891019-0037

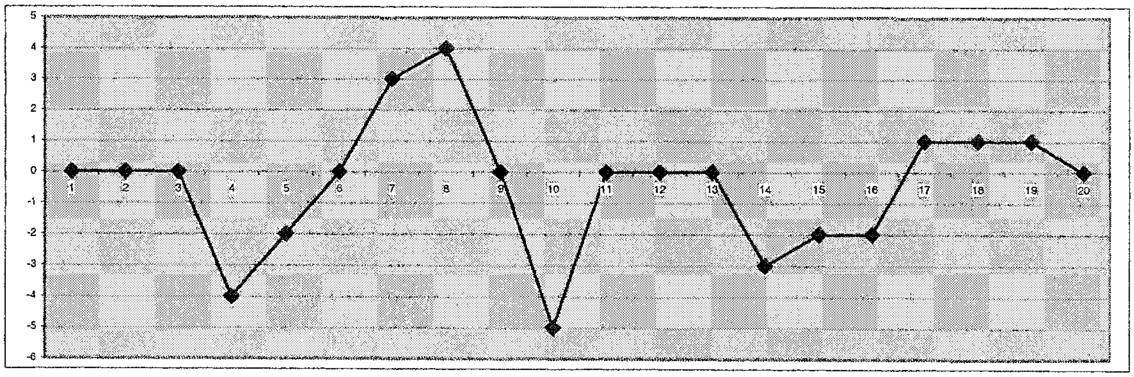

Figure A45: d098e\AP900421-0075

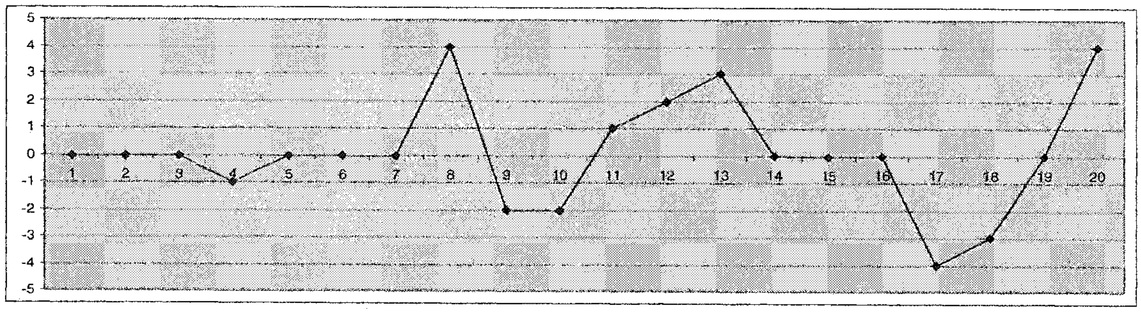

Figure A46: d098e\AP900424-0048

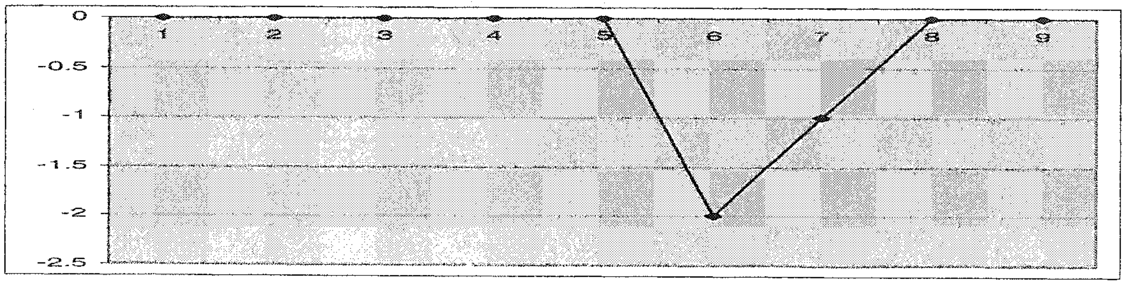

Figure A47: d098e\AP900424-0081

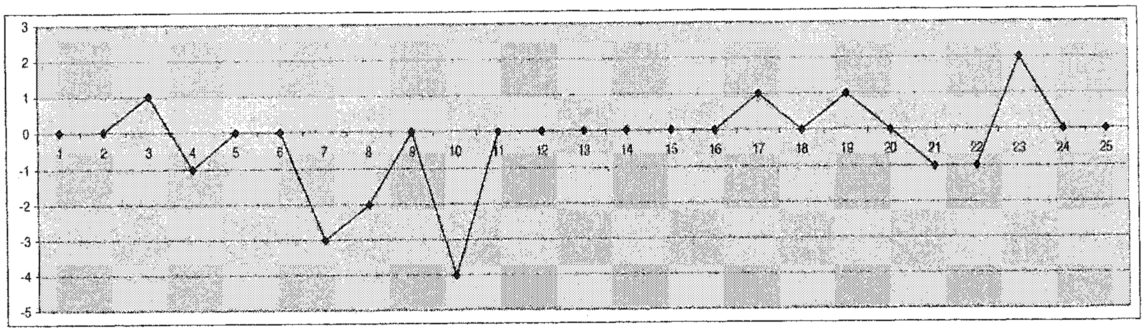

Figure A48: d098e\AP900424-0096

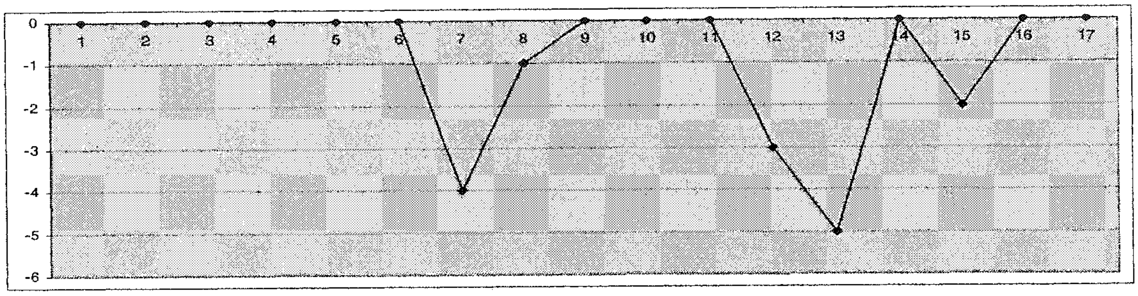

Figure A49: d098e\AP900425-0013

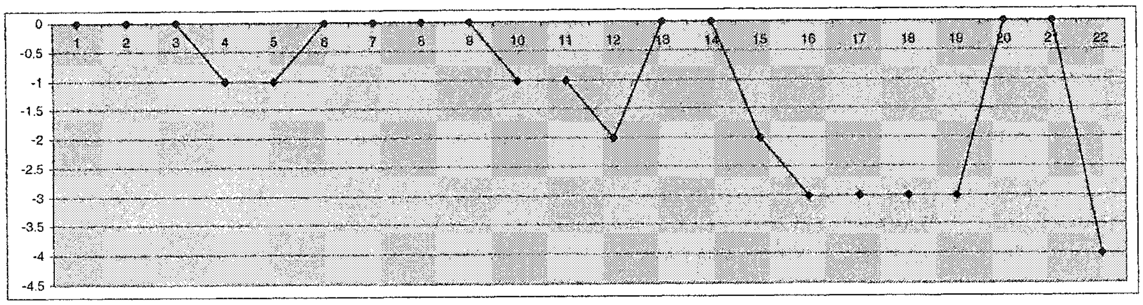

Figure A50: d098e\AP900425-0146

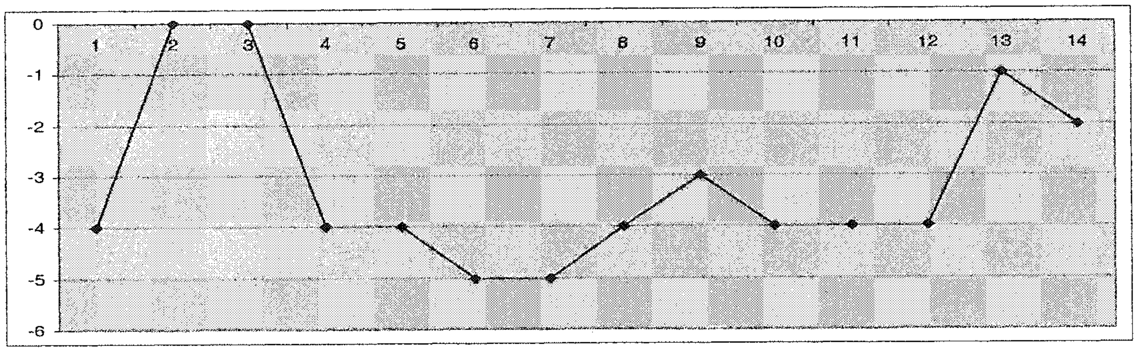

Figure A51: d100e\AP880808-0040

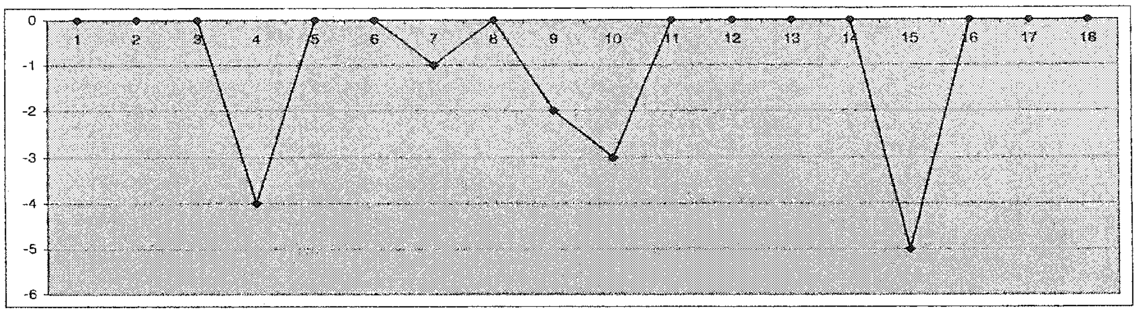

Figure A52: d100e\AP880915-0066

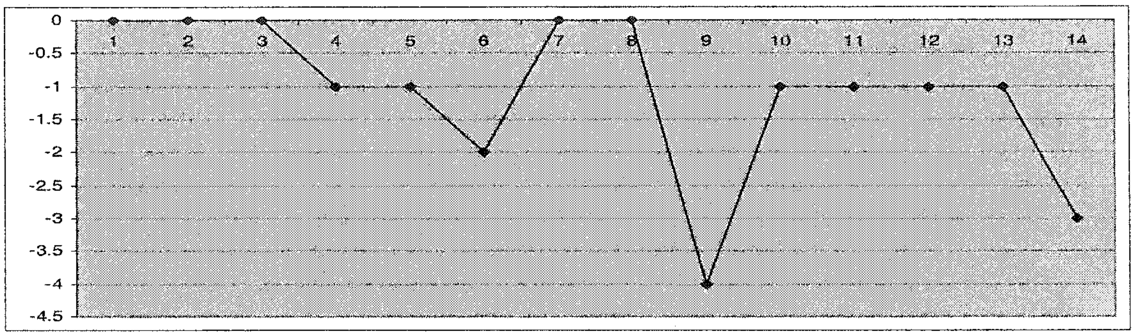

Figure A53: d100e\AP881001-0104

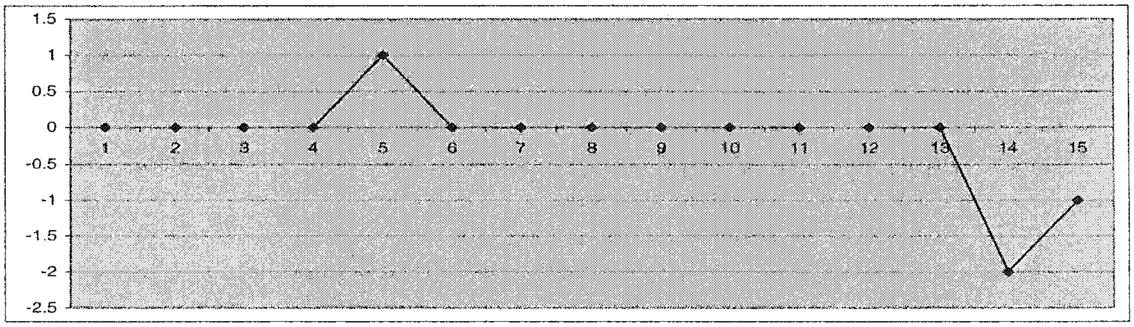

Figure A54: d100e\AP900505-0127

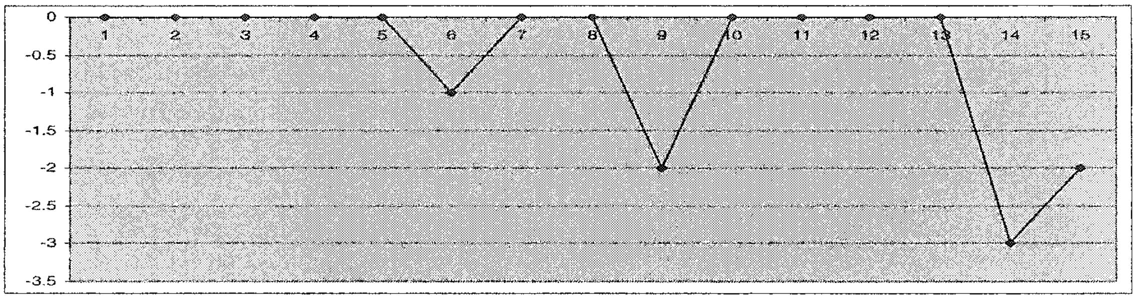

Figure A55: d100e\AP901009-0072

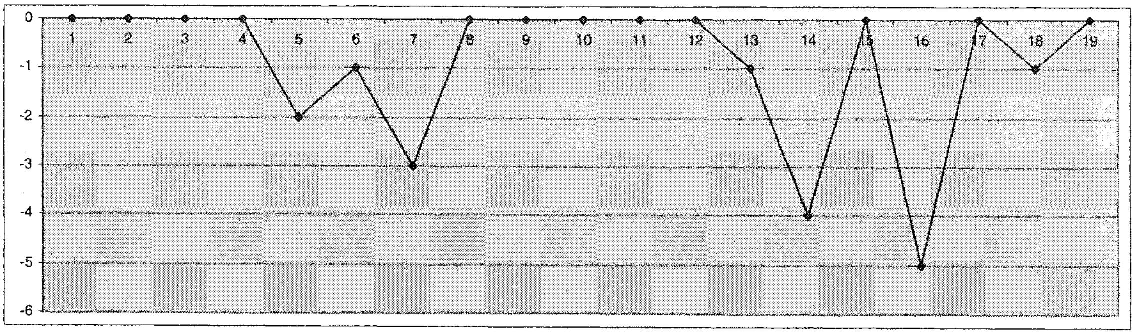

Figure A56: d105g\AP880729-0155

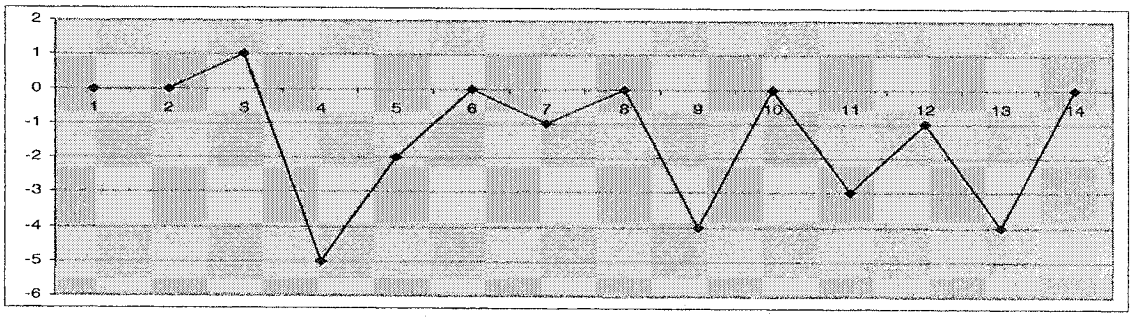

Figure A57: d105g\AP891213-0164

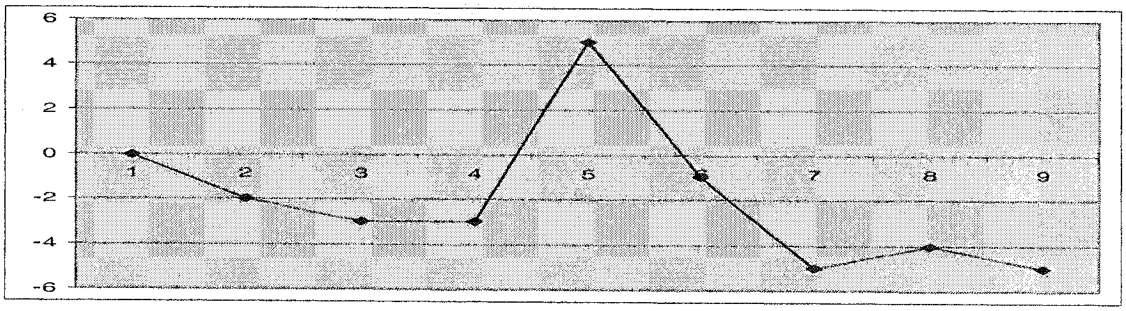

Figure A58: d105g\AP900613-0195

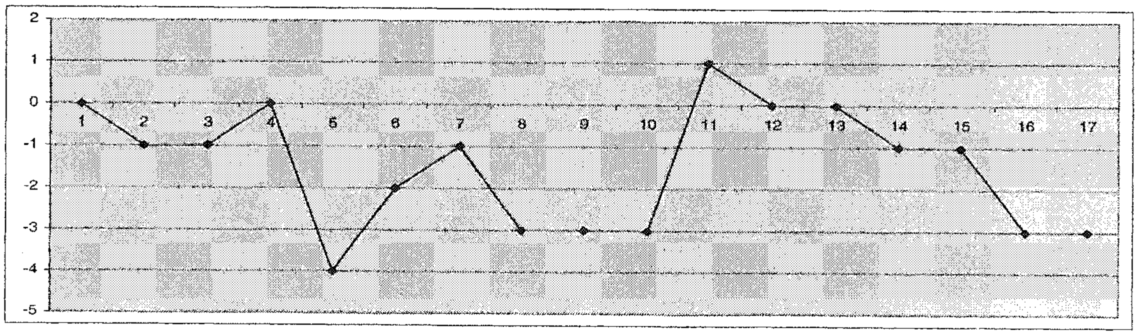

Figure A59: d105g\AP901230-0022

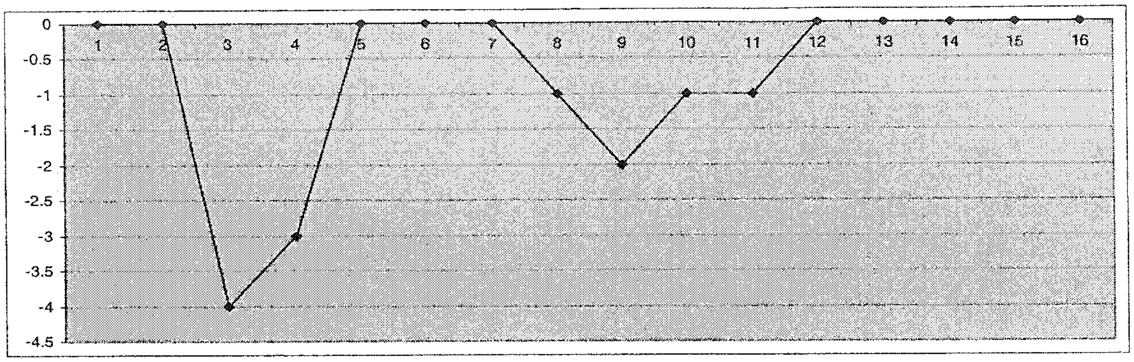

Figure A60: d108g\AP900807-0029

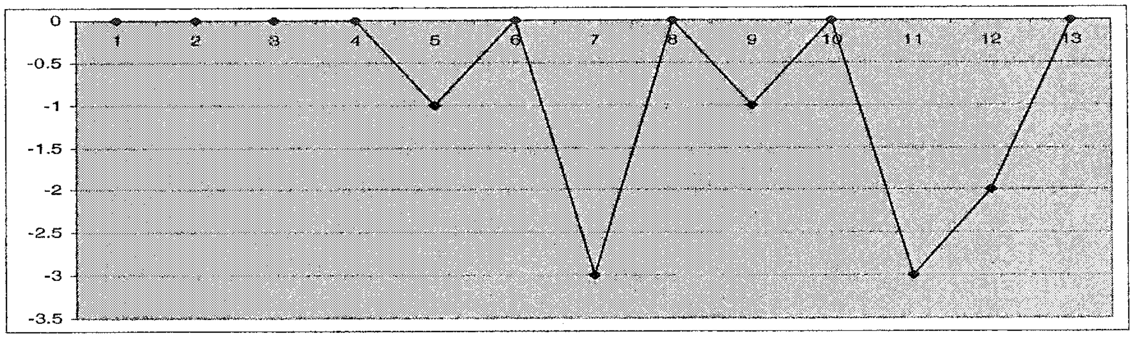

Figure A61: d108g\AP900807-0093

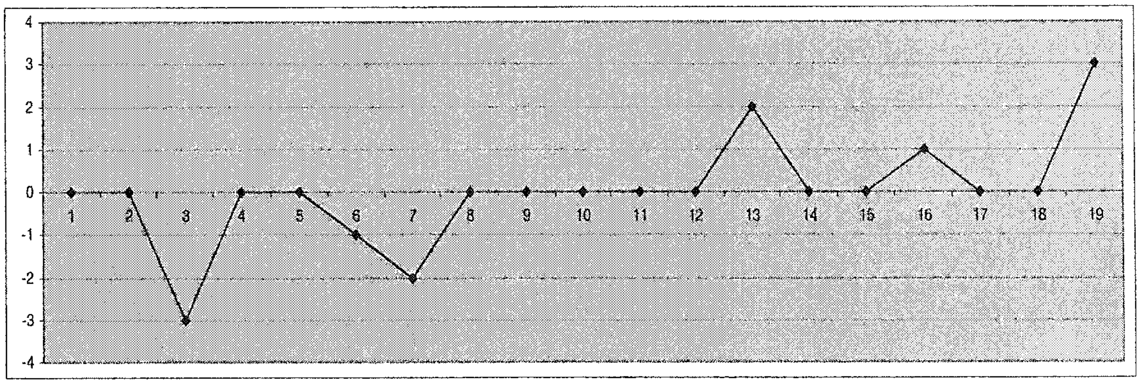

Figure A62: d108g\AP900808-0030