PRC Inc:
DESCRIPTION OF THE PAKTUS SYSTEM USED FOR MUC-3

Bruce Loatman
PRC Inc
1500 Planning Research Drive
McLean, VA 22102
loatman_bruce@po.gis.prc.com

BACKGROUND

The PRC Adaptive Knowledge-based Text Understanding System (PAKTUS) has been under development as an Independent Research and Development project at PRC since 1984. The objective is a generic system of tools, including a core English lexicon, grammar, and concept representations, for building natural language processing (NLP) systems for text understanding. Systems built with PAKTUS are intended to generate input to knowledge based systems or data base systems. Input to the NLP system is typically derived from an existing electronic message stream, such as a news wire. PAKTUS supports the adaptation of the generic core to a variety of domains: JINTACCS messages, RAINFORM messages, news reports about a specific type of event, such as financial transfers or terrorist acts, etc., by acquiring sublanguage and domain-specific grammar, words, conceptual mappings, and discourse patterns. The long-term goal is a system that can support the processing of relatively long discourses in domains that are fairly broad with a high rate of success.

APPROACH

PAKTUS may be viewed from two perspectives. In one view it is seen as a generic environment for building NLP systems, incorporating modules for lexical acquisition, grammar building, and conceptual template specification. The other perspective focuses on the grammar, lexicon, concept templates, and parser already embedded within it, and views it as an NLP system itself. The early emphasis in developing PAKTUS was on those components supporting the former view. The grammar and lexicon that form the common core of English, as well as the stock of generic conceptual templates, entered PAKTUS primarily as a side effect of the testing of extensions to the NLP system development environment. More recent work has focused on extending the linguistic knowledge within the overall architecture, such as prepositional phrase attachment, compound nominals, temporal analysis, and metaphorical usage, and on adapting the core to particular domains, such as RAINFORM messages or news reports.

The first step in this project was an evaluation of existing techniques for NLP, as of 1984. This evaluation included implementing rapid prototypes using techniques as in [1], [2], [3], and [4]. Judging that no one technique was adequate for a full treatment of the NLP problem, we adopted a hybrid approach, breaking the text understanding process into specialized modules for text stream preprocessing, lexical analysis, including morphology, syntactic analysis of clauses, conceptual analysis, domain-specific pattern matching based on an entire discourse (e.g., a news report), and final output-record generation.

Knowledge about word morphology was drawn from [5] and is represented as a semantic network, as is lexical and semantic knowledge in general. The grammar specification has been based on our analysis of message text, and draws from [5], [6], and [7]. It was first implemented as an augmented transition network (ATN), using a linguistic notation similar to that in [4]. This implementation relies on an interactive graphic interface to specify and debug grammar rules. More recent investigations focus on alternative formalisms.

Our conceptual analysis combines aspects of conceptual dependency [1], [8], case grammar [9], semantic preferences [10], and psychology [11]. We provided a feedback loop from the conceptual analyzer to the syntactic parser for faster, more accurate analysis. We have found empirically that the current parser usually runs in linear time (on a Sun 3/260, about 0.1 second per word, regardless of sentence length). This is a result of the feedback together with "look ahead" tests at critical decision points. Those infrequent sentences requiring more time are terminated by the parser, which returns the longest parsed substring, or a run-on analysis. The resultant loss of recall is more than compensated for in increased system throughput. MUC-3 corpus sentences that PAKTUS parses completely (about 47% of the total corpus) average about two seconds of parse time, compared to about ten seconds each for partial and run-on parses (about 51% of the total corpus).

Our first version of the domain-specific discourse pattern matcher was based on [2], but a more versatile version, based on specification-by-examples, was added during MUCK 2 development. This uses clause, sentence, and noun

phrase semantic and syntactic patterns, and was used again in MUC-3. We have begun implementing discourse-level pattern matching (somewhat like scripts), but this was not sufficiently developed for use in MUC-3.

In addition to these functional NLP components, PAKTUS has a broad set of development tools, including grammar construction tools and debugger, a lexical acquisition interface, conceptual specification tools, domain pattern builders, and some automatic learning capabilities. These greatly facilitated adaptation of the system to the MUC-3 task.

System Architecture

PAKTUS integrates a variety of methods that have had some success in the past. The architecture of the PAKTUS-based NLP system used for MUC-3 is shown in Figure 1.

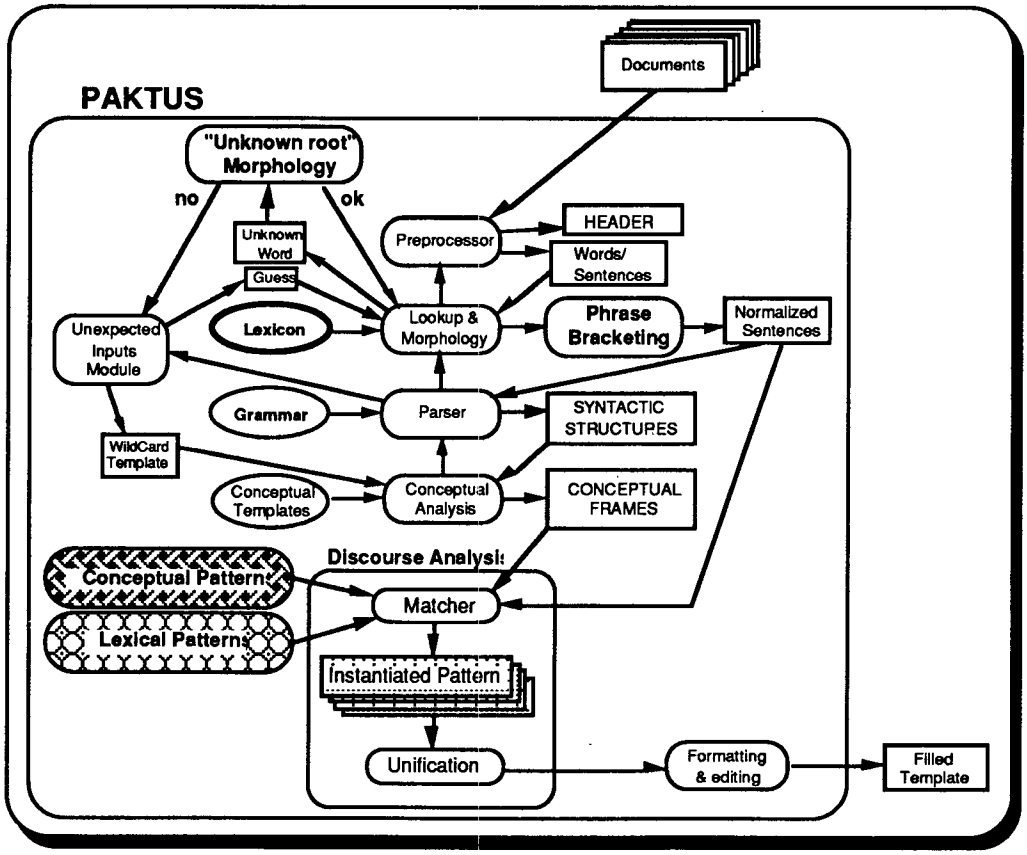

Figure 1: PAKTUS Architecture used for MUC-3.

Processing begins with the arrival of an electronic stream of text, such as the MUC-3 corpus. The first function performed is the decomposition of the stream of characters into individual messages, message segments, sentences, and strings of words, based on document format specifications contained in a MUC-3 document specification template. The words identified by the preprocessor are mapped into entries in the lexicon which contain information about their syntax and semantics, as illustrated in Figure 2 for the word "knows". The lexicon, contained in five databases, contains separate information for root words ("words" in Figure 2), concepts, and surface forms or "tokens". The latter are mapped into data structures based on the roots. These mappings are contained the "parses" database of Figure 2. Compound words and idioms are first mapped into synthetic tokens, and then processed like other surface forms. All this information is organized in memory in two networks: a lexical net and a concept net. These two networks are linked by conceptual associations, as illustrated in Figure 2. The words, concepts, and associations are brought into memory only as needed in processing text. PAKTUS includes an object-oriented DBMS that performs these lexicon operations [12].

When words are encountered that have never been seen previously by PAKTUS, it tries to analyze these morphologically. The morphology module has information about approximately 250 affixes, one of which, -en, is illustrated in Figure 3. It analyzes words in terms of known roots and the affixes, although some words can be adequately analyzed without any knowledge of the root (e.g., any word not in the lexicon that ends in "ology"

denotes an "information domain"). It derives both syntactic information and semantic information, producing an internal PAKTUS representation. In addition, for MUC-3, we added morphological heuristics for guessing syntactic and semantic information in many cases, even when the root is unknown (e.g., an unrecognized word ending in "ation" might be an abstract noun). If all of this fails to identify the word, the system deduces as much as it can from the syntactic and semantic context during parsing.

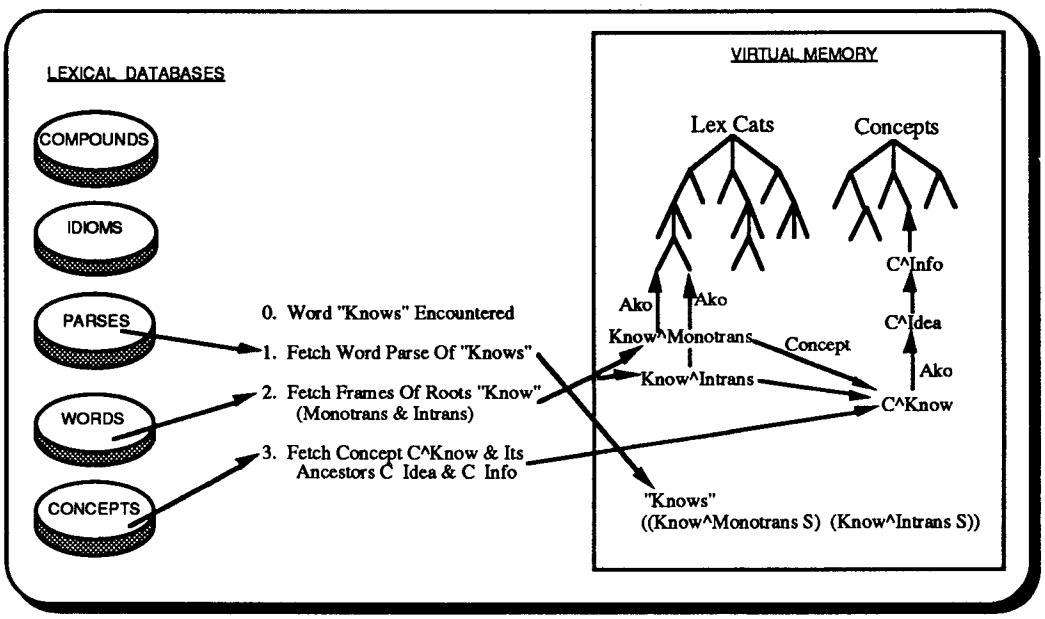

Figure 2: Structure and Operation of PAKTUS Lexicon.

```
(S^EN
  (AKO (JSUFF))
  (APPLIES-TO (L^ADJ))
  (LEX-CHANGE
    ((L^ADJ L^MONOTRANS
            L^INTRANS)))
  (ROOT-PREP (S^R72))
  (LEX ("EN"))
  (CON-MAP
    ((L^INTRANS  C^BECOME-ROOT)
     (L^MONOTRANS  C^CAUSE-BE-ROOT
```

Figure 3: Morphological information associated with the suffix "en" in PAKTUS.

The next step in processing the text is to parse the sentences syntactically, according to a grammar specification, to generate syntactic configurations. The conceptual analyzer then maps the syntactic configurations into conceptual frames (concept structures with roles filled by phrase constituents), usually resolving much ambiguity in the process. If the syntax cannot be mapped into any conceptual frame, it is rejected and the syntactic parser tries alternatives. The first two levels of the conceptual network are shown in Figure 4. The conceptual roles used in PAKTUS are shown in Figure 5.

The discourse analyzer collects all the conceptual frames for an entire narrative (i.e., an entire news report for the MUC-3 application) and produces application-specific structures that represent the information that is to be extracted from the document, based on the discourse template specifications. These structures are then reformatted into simulated MUC-3 data base updates (i.e., filled templates). An example is given below.

There are important feedback points in this process, as shown in Figure 1. For example, the conceptual analyzer may notify the syntactic parser that a proposed parse is semantically unacceptable, signalling that an alternative parse should be attempted. This semantic testing is always invoked at the clause level, and sometimes sooner. In addition, when confronted with two computationally expensive paths, "look ahead" procedures that quickly scan the sentence are invoked to decide which to try first. For example, a past participle following a noun

may or may not signal the beginning of a relative clause in which the noun is the direct object. In this case, a partial conceptual analysis quickly determines whether the noun can be mapped into any concept associated with the verb. If it cannot, the relative clause path is not pursued.

```
(C^PRIMITIVE
  (C^BOUND
    C^BEGIN C^END C^EVER C^WHOLE)
  (C^CHANGE
    C^AID C^DAMAGE C^VARY C^LAW C^LEAD C^POWER C^SUBJUG
    C^GO C^EVENT C^NEW C^TIME C^REST C^WEAK)
  (C^LIFE
    C^BODY C^ANIMAL C^BUG C^LIVE C^FOOD C^SIP C^VEGT
    C^SOMA C^GARB C^HOME C^HOLY C^KIN C^LADY C^YOUNG
    C^CRIME C^SICK C^EDUCATE C^MALE C^MEDIC
    C^MART C^MONEY C^WORK C^PLAY C^SEX)
  (C^ARTIFACT
    C^FURN C^MECHAN C^VEHIC)
  (C^STUFF
    C^FABRIC C^MATERIAL C^VAPOR C^ION C^EARTH)
  (C^INTER-REL
    C^PROX C^REL-LOC C^HOLLOW C^JOIN
    C^HAVE C^LACK C^GROUP C^SEPAR C^WE)
  (C^SELF-REL
    C^PURITY C^AMBIENCE C^PSYCH C^GOODNESS)
  (C^INFO
    C^HEAR C^VIEW C^IDEA C^MEASURE
    C^TALK C^WRITE C^MUSIC C^COMPUTE C^TRUTH)
  (C^GRADE
    C^MUCH C^LARGE C^LITTLE C^QUICK C^SOME C^EASY C^SOLE)
  (C^PHYSIC
    C^BULGE C^FLOW C^PATH C^PLACE C^ASTRON)
  C^UNKNOWN)
```

Figure 4: First two levels of the PAKTUS concept network.

```
(Role
  (Prop-role
    (R^Doer
      R^Agent R^Instr)
    (R^Object
      R^Affected R^Experiencer R^Companion
      (R^Effect
        R^Result R^Event)
      R^Proposition R^Focus R^Purpose R^Material
      (R^Resistance
        R^Opponent R^Opposition))
    R^Attribute
    (R^Loc
      R^Owner
      (R^Source
        R^Origin R^Donor)
      (R^Goal
        R^Dest R^Recipient)
      R^Extent R^Place
      (R^Route
        R^Path R^Medium))
    R^Co-con)

  (Modal-role
    (R^Time
      R^When R^Begin R^Finish R^Duration R^When-adv)
    R^Manner R^Beneficiary R^Method R^Cause R^Accomplice
    R^Possessed R^Precondition R^Tool R^Place R^Purpose)

  Super-con)
```

Figure 5: Conceptual roles used in PAKTUS

APPLICATION TO MUC-3

To apply PAKTUS to MUC-3, five tasks were performed. Due to the modular design of PAKTUS, these are clearly delineated and were performed by different people. Some tasks must be initiated in sequence, but may be cascaded as the corpus of text is processed, so that, except for a brief period at the beginning of the task, work proceeded in parallel. The number of changes to various knowledge bases is summarized in Figure 6. The five tasks were:

- Build a template specifying the formats of the input streams (dev-muc3, tst1 and tst2). This was easy and required about a day to perform.
- Read in the documents and update the lexicon using the PAKTUS interactive graphic interface. This was relatively easy for those words (typically nouns) that only require categorization, but not conceptual mapping specifications (as verbs do). The latter has often been done successfully by relying on PAKTUS default values.
- Adapt the grammar to the sublanguage of the application. Actually changing the grammar is easy with the PAKTUS interactive graphic tools for this, but determining what is the grammar of the sublanguage may be quite difficult, requiring much linguistic knowledge and study of the corpus. Changes for MUC-3 were minor.
- Define the application-specific discourse templates. This is the least developed component of PAKTUS, and the one that will receive the most attention in continuing work, such as for MUC-4. For MUC-3, phrase and sentence-level patterns were defined. A function unified these and mapped them into the 18-slot MUC-3 templates.
- Specify and implement the interface to the application system (the MUC-3 template fills). This was tedious, but easy compared to the other tasks. It is strictly conventional software engineering.

	Before MUC-3	Added for MUC-3
Words (Stems)	5448	3565
Tokens	8552	4307
Compounds	102	426
Idioms	45	22
Concepts	386	0
Subconcepts	15	0
Verb categories	16	0
Nominal categories	398	8
Adverb categories	10	0
Closed categories	41	0

Figure 6: Knowledge structures added to PAKTUS for MUC-3

Example of MUC-3 Document Processing

Message number 99 from the "test1" set, which is reprinted in Appendix H, will be used to illustrate PAKTUS's operation for MUC-3. PAKTUS processes text sequentially, first stripping off the document header, then identifying sentences, which are processed syntactico-semantically one at a time, after which all the results are passed to the discourse component.

Figure 7 shows the raw, unprocessed text of the first sentence, followed by the preprocessed sentence, in which word boundaries have been identified. Note that "Soviet Union" is treated as a single word, since it names an entity represented in the lexicon.

The lexical analysis of this sentence is shown in Figure 8. Each word has one or more senses, represented as a root symbol, which is generally the concatenation of the English token, the "^" character, and the PAKTUS lexical category (e.g., "Report^Monotrans"), or as a simple structure involving a root, lexical category, inflectional mark, and sometimes a conceptual derivation (e.g. the structure "(Report^Monotrans L^Effect-mark Base C^It-got)" represents the adjective sense of "reported"). For each word, all senses in the PAKTUS lexicon are fetched or derived at this time; disambiguation is generally delayed until the syntactic and semantic phases. An exception in this

example is the word "tonight" which has been replaced by the date from the dateline of this MUC-3 news report. The syntactic and conceptual analyses of this sentence are shown in Figures 9 and 10, respectively. Note that conceptual structures are produced for some nouns (e.g., "embassies"), not just for verbs.

```
*** raw sentence:
POLICE HAVE REPORTED THAT TERRORISTS TONIGHT BOMBED THE
EMBASSIES OF THE PRC AND THE SOVIET UNION.

*** preprocessed words:
("POLICE" "HAVE" "REPORTED" "THAT" "TERRORISTS" "TONIGHT"
 "BOMBED" "THE" "EMBASSIES" "OF" "THE" "PRC"
 "AND" "THE" "SOVIET UNION")
```

Figure 7: Result of preprocessing the first sentence of test1 document number 99.

Figures 11 and 12 show the syntactic and conceptual analysis of the fifth sentence, which contains a difficult conjunction that PAKTUS handled correctly. Figure 13 shows one of the templates PAKTUS generated for this news report. It is fairly representative of our work focus in MUC-3, in that some slots are filled in correctly, some incorrectly (e.g., location), and some ignored (e.g., perpetrators), due to the limits of our development in this application.

```
POLICE HAVE REPORTED THAT TERRORISTS TONIGHT BOMBED THE
EMBASSIES OF THE PRC AND THE SOVIET UNION.

*** lexical analysis:
((Police^Specialist)
 (Have1^Monotrans Have^Monotrans Have^Have
   Have^Intrans Have2^Intrans)
 ((Report^Monotrans L^Effect-mark Base C^It-got)
   (Report^Intrans L^Intrans S^Ed)
   (Report^Monotrans L^Monotrans S^Ed))
 (That^Binder That^Rel That^Far)
 ((Terror^Emotion L^User S^S C^Uses))
 (("25-oct-1989" L^Time-date Base))
 ((Bomb^Monotrans L^Effect-mark Base C^It-got)
   (Bomb^Monotrans L^Monotrans S^Ed)
   (Bomb^Intrans L^Intrans S^Ed))
 (The^Det)
 ((Embassy^Building L^Building S^S)
   (Embassy^Govt-org L^Govt-org S^S))
 (Of^Particle Of^Prep)
 (The^Det)  (China^Nation)
 (And^Conj)  (The^Det)  (Russia^Nation))
```

Figure 8: Lexical analysis of the first sentence of test1 document number 99.

```
POLICE HAVE REPORTED THAT TERRORISTS TONIGHT BOMBED THE
EMBASSIES OF THE PRC AND THE SOVIET UNION.

*** COMPLETE parse:
Syntax:
(S ("Main-verb11" (Report^Monotrans L^Monotrans S^Ed))
   ("Subject09"
      (Np  ("Head10" Police^Specialist)))
   ("Prop93"
    (T^
      ("Main-verb07" (Bomb^Monotrans L^Monotrans S^Ed))
      ("Subject05"
        (Np  ("Head06" (Terror^Emotion L^User S^S C^Uses))))
      ("Do95"
        (Np  ("Head03" (Embassy^Building L^Building S^S))
             ("Det04" The^Det)))
      ("Adv94" ("25-Oct-1989" L^Time-date Base)))))
```

Figure 9: Syntactic analysis of the first sentence.

Figure 14 illustrates the ability of PAKTUS to deal with unknown words, which is essential in any application that continually processes new text. This shows the syntactic analysis of a sentence from one of the "test2" reports. It contains three words that can not be derived from the PAKTUS lexicon: "Estevez", "MPTL", and "supposed". PAKTUS made assumptions about each word, based on morphology and syntactico-semantic context. It was able to produce a reasonably accurate parse, by guessing that "Estevez" is a Spanish name, and recognizing that "MPTL" is a noun in apposition with the preceding noun phrase, and that "supposed" must in this case be a passive voice monotransitive verb.

```
POLICE HAVE REPORTED THAT TERRORISTS TONIGHT BOMBED THE
EMBASSIES OF THE PRC AND THE SOVIET UNION.

Conceptual frame:
(C^Assert ("R^Agent09" (F409 ("Head10" Police^Specialist)))
 ("R^Proposition93"
   (C^Damage ("R^Agent05" (F405
                ("Head06" (Terror^Emotion L^User S^S C^Uses))))
     ("R^Affected95"
       (C^Attend ("Head03" (Embassy^Building L^Building S^S))
         ("R^Focus96"
           (F396 ("Head01" China^Nation)
           ("Conj97" (F397 ("Head98" Russia^Nation)))
           ("Conjoiner00" And^Conj)))
       ("R^Place95" "@F395") ("R^Place95" "@F395")))
     ("R^Opponent95"
       (C^Attend ("Head03" (Embassy^Building L^Building S^S))
         ("R^Focus96"
           (F396 ("Head01" China^Nation)
           ("Conj97" (F397 ("Head98" Russia^Nation)))
           ("Conjoiner00" And^Conj)))
       ("R^Place95" "@F395") ("R^Place95" "@F395")))
     ("R^When-adv94" ("25-Oct-1989" L^Time-date Base)))))
```

Figure 10: Conceptual analysis of the first sentence.

```
POLICE SAID THE ATTACKS WERE CARRIED OUT ALMOST
SIMULTANEOUSLY AND THAT THE BOMBS BROKE WINDOWS
AND DESTROYED THE TWO VEHICLES.

Syntax:

(S ("Main-verb28" (Say^To-io L^To-io S^Ed))
 ("Subject26" (Np ("Head27" Police^Specialist)))
 ("Prop97"
  (T1 ("Main-verb03" (Carryout^Monotrans L^Monotrans S^Ed))
   ("Subject01" (Np ("Head02" Someone^Some)))
   ("Do98" (Np ("Head99" (Attack^Task L^Task S^S)) ("Det00" The^Det)))
   (" Conj04"
    (Conj^ ("Main-verb11" ( Break^Intrans L^Monotrans Past C^Cause-it) )
     ("Subject08" (Np ("Head09" ( Bomb^Bomb L^Bomb S^S ))
                   ("Det10" The^Det)))
     ("Do06" (Np ("Head07" ( Window^Struct-part L^Struct-part S^S))))
     (" Conj12"
       (Conj^ ("Main-verb20" ( Destroy^Monotrans L^Monotrans S^Ed))
        ("Subject08" (Np ("Head09" ( Bomb^Bomb L^Bomb S^S ))
                      ("Det10" The^Det)))
        ("Do14" (Np
          ("Head16" ( Vehicle^Transport-system L^Transport-system S^S ))
          ("Det19" The^Det) ("Numeral15" ( 2 L^Cardinal Base)))))))))
   ("Adv22" (Simultaneous^Time-rel L^Manner Base C^Do-like))
   ("Adv23" Almost^Intensity))))
```

Figure 11: Syntactic analysis of the fifth sentence.

POLICE SAID THE ATTACKS WERE CARRIED OUT ALMOST SIMULTANEOUSLY AND THAT THE BOMBS BROKE WINDOWS AND DESTROYED THE TWO VEHICLES.

Conceptual frame:

```
(C^Assert ("R^Agent26" (F526 ("Head27" Police^Specialist)))
 ("R^Proposition97"
  (C^Act  ("R^Agent01" (F501 ("Head02" Someone^Some)))
   ("R^Result98" (F498 ("Head99" (Attack^Task L^Task S^S))))
   ("Conj04"
    (C^Cause-it
     ("R^Instr08"
        (C^Damage ("Head09" (Bomb^Bomb L^Bomb S^S))
                  ("R^Instr08" "@F508")))
     ("R^Event05"
      (C^Deform
         ("R^Affected06" (F506
            ("Head07" (Window^Struct-part L^Struct-part S^S))))))
     ("Conj12"
      (C^Damage
        ("R^Instr08"
          (C^Damage  ("Head09" (Bomb^Bomb L^Bomb S^S))
                     ("R^Instr08" "@F508")))
        ("R^Affected14"  (F514
          ("Head16" (Vehicle^Transport-system  S^S)))))))))))
```

Figure 12: Conceptual analysis of the fifth sentence.

```
0. MESSAGE ID                    TST1-MUC3-0099
1. TEMPLATE ID                   4
2. DATE OF INCIDENT                - 25 OCT 89
3. TYPE OF INCIDENT              BOMBING
4. CATEGORY OF INCIDENT          TERRORIST ACT
5. PERPETRATOR: ID OF INDIV(S)   "TERRORISTS"
6. PERPETRATOR: ID OF ORG(S)     -
7. PERPETRATOR: CONFIDENCE       REPORTED AS FACT: -
8. PHYSICAL TARGET: ID(S)        "EMBASSIES"
9. PHYSICAL TARGET: TOTAL NUM    PLURAL
10. PHYSICAL TARGET: TYPE(S)     DIPLOMAT OFFICE OR
                                 RESIDENCE: "EMBASSIES"
11. HUMAN TARGET: ID(S)            -
12. HUMAN TARGET: TOTAL NUM        -
13. HUMAN TARGET: TYPE(S)          -
14. TARGET: FOREIGN NATION(S)      -
15. INSTRUMENT: TYPE(S)          *
16. LOCATION OF INCIDENT         CHINA
17. EFFECT ON PHYSICAL TARGET(S)   SOME DAMAGE: "EMBASSIES"
18. EFFECT ON HUMAN TARGET(S)    INJURY: -
```

Figure 13: Sample filled-template for message 99.

```
JAIME ESTEVEZ, A MEMBER OF THE BREAD, LAND, WORK, AND
FREEDOM MOVEMENT [MPTL] WAS TAKEN TO THE USULUTAN
JAIL TO BE MURDERED AND THE CRIME WAS SUPPOSED TO GO
UNNOTICED.

Syntax:

(S ("Main-verb64" (Take2^Monotrans L^Monotrans Past-part))
 ("Subject62" (Np ("Head63" Someone^Some)))
 ("Do39"
  (Np ("Head59" ("Estevez" L^Espn-person Base C^Unknown))
   ("Desc61" Jaime^Male)
   (" App40"
    (Np ("Head57" Bread^Food) ("Det58" The^Det)
     ("Conj42"  (Np ("Head55" Land^Geographical-area)
      ("Conj43"  (Np ("Head53" Work^Activity)
       ("Conj44"  (Np ("Head50" (Move^Intrans L^Abstract Base C^Act-of))
        ("Desc46" (Free^Subst-compar L^Abstract Base C^Situation-of))
        ("  App47" (Np ("Head48" ("MPTL" L^Common Base C^Unknown ))))))))))
    ("Quant41" Member^Body-part)))
 ("Mods78"  (Pp ("Prep79" To^Prep)
  ("Prep-obj26"
   (Np ("Head35" Jail^Dwelling) ("Det38" The^Det)
       ("Desc37" Usulutan^City-or-town)
   ("Prop27"
     (Z^ ("Main-verb32" (Murder^Monotrans L^Monotrans S^Ed))
         ("Subject30" (Np ("Head31" Someone^Some)))
         ("Do28" (Np ("Head29" Someone^Some)))))))))
 ("Conj65"
  (Conj^ ("Main-verb76" (" Suppos" L^Monotrans S^Ed C^Unknown))
   ("Subject69" (Np ("Head70" Someone^Some)))
   ("Do73" (Np ("Head74" Crime^Activity) ("Det75" The^Det)))
   ("Aps66"
    (Z^ ("Main-verb71" Go^Copula) ("Subject69" (Np ("Head70" Someone^Some)))
     ("Comp67"
      (Np ("Head68" (Notice^Monotrans L^Effect-mark Base C^Not C^It-got)))))))))
```

Figure 14: Sample sentence with unknown words, as parsed by PAKTUS.

REFERENCES

[1] Schank, R, *Conceptual Information Processing*, New York: North Holland, 1975.

[2] Dyer, M, *In-Depth Understanding*, Cambridge, MA: MIT Press, 1983.

[3] Marcus, M, *Theory of Syntactic Recognition for Natural Language*, Cambridge, MA: MIT Press, 1980.

[4] Winograd, T, *Language as a Cognitive Process*. Reading, MA: Addison-Wesley, 1983.

[5] Quirk, R, Greenbaum, S, Leech, G, and Svartvik, J, *A Comprehensive Grammar of the English Language*, New York: Seminar Press, 1985.

[6] Jespersen, O, *Essentials of English Grammar*, University, AL: University of Alabama Press, 1964.

[7] Sager, N, Friedman, C, and Lyman, M, *Medical Language Processing: Computer Management of Narrative Data*, Reading: Addison-Wesley 1987.

[8] Lebowitz, M, "Memory-Based Parsing", *Artificial Intelligence*, vol. 21, pp. 363-404, 1983.

[9] Cook, W, *Case Grammar: Development of the Matrix Model*, Washington DC: Georgetown University Press, 1979.

[10] Wilks, Y, Huang, X, and Fass, D, "Syntax, Preference and Right Attachment", *Proceedings of the Ninth IJCAI*, 1985.

[11] Laffal, J, *A Concept Dictionary of English*, Essex, CT: Gallery Press, 1973.

[12] Loatman, B, and Levesque, R, "A Portable Object-Oriented DBMS", *Proceedings of Database Colloquium '91*, San Diego: AFCEA, 1991.