BBN: Description of the PLUM System as Used for MUC-6

BBN Systems and Technologies
70 Fawcett Street
Cambridge, MA 02138
weischedel@bbn.com

ABSTRACT

This paper provides a quick summary of our technical approach, which has been developing since 1991 and was first fielded in MUC-3. First a quick review of what is new is provided, then a walkthrough of system components. Perhaps most interesting is our analysis, following the walkthrough, of what we learned through MUC-6 and of what directions we would take now to break the performance barriers of current information extraction technology.

TECHNICAL APPROACH

Our approach is a synthesis of statistical and linguistic models of language, embodied for MUC-6 in the PLUM system (Probabilistic Language Understanding Model). We began this research in 1991 and applied it that year in MUC-3. Since that time, we have successfully applied probabilistic models to the following:

- part of speech tagging, using Hidden Markov Models (Weischedel, et al., 1993),
- judgments of relevance of text (at the paragraph level) via a log odds model (Ayuso, et al., 1992),
- learning semantic information (selection restrictions, or case frames) from supervised training (Weischedel, et. al., 1991, 1993),
- learning word and word group associations (Matsukawa, 1993),
- an example-based correction technique for segmentation and part-of-speech labeling for Japanese (Matsukawa, Miller, and Weischedel, 1993), and
- induction of a statistical grammar and parser from training data (Magerman, 1995).

Not all techniques are necessarily used in each domain, nor necessarily effective in each domain. In MUC-6, there was so little data that only domain-independent techniques were employed. More linguistically motivated knowledge bases are used in almost every domain, such as case frames for verbs and nouns, verb complementizer structure, etc.

SUMMARY OF WHAT'S NEW

In the last two years we have ported part or all of the PLUM system to several new languages (Chinese, German, Japanese, and Spanish) and new domains (law enforcement, name finding, heterogeneous newswire sources, and labor negotiations). Though we have a new, fully trainable, full parser of English (Magerman, 1995), there was insufficient time to integrate it into PLUM for the evaluation; as an independent component, it appears to have achieved the highest published evaluation scores for parsers.

The new software developments employed in MUC-6 are

- a stand-alone, C-based name spotter (IdentiFinder™), a rewrite of the initial components of PLUM,
- a more robust message reader,
- a revised discourse component and output generator to more cleanly separate discourse structures from the final template structure, and
- a semantic inference component.

We have begun making a distinction between *lightweight* techniques and *heavyweight* processing. IdentiFinder is made up solely of lightweight techniques, i.e., those that rely only on local processing, do not involve deep understanding, and can be optimized. The lightweight procedures in IdentiFinder are SGML recognition, hidden Markov models, finite state pattern recognition, and SGML output.

By heavyweight processing, we mean procedures that depend on global evidence and involve deeper understanding. The SPATTER full parser of English and the new semantic inference procedure are examples.

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE 1995	2. REPORT TYPE	3. DATES COVERED 00-00-1995 to 00-00-1995
4. TITLE AND SUBTITLE **BBN: Description of the PLUM System as Used for MUC-6**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **BBN Technologies,10 Moulton Street,Cambridge,MA,02238**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**15**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

KEY SYSTEM FEATURES

Two key design features of PLUM are: statistical language modeling with the associated learning algorithms, and partial understanding. The first key feature is the use of *statistical modeling to guide processing*. For the version of PLUM used in MUC-6, part of speech information was determined by using well-known Markov modeling techniques embodied in BBN's part-of-speech tagger POST (Weischedel, et al., 1993).

A second key feature is *partial understanding*, by which we mean that all components of PLUM are designed to operate on partially interpretable input, taking advantage of information when available and not failing when it is unavailable. Neither a complete grammatical analysis nor complete semantic interpretation is required. The system pieces together a model of the whole from the parts of the text it can understand.

PLUM PROCESSING STAGES

The PLUM architecture is presented in Figures 1 and 2. Ovals represent declarative knowledge bases; rectangles represent processing modules. A more detailed description of the system components, their individual outputs, and their knowledge bases is presented in Ayuso et al., 1993. The processing modules are briefly described below.
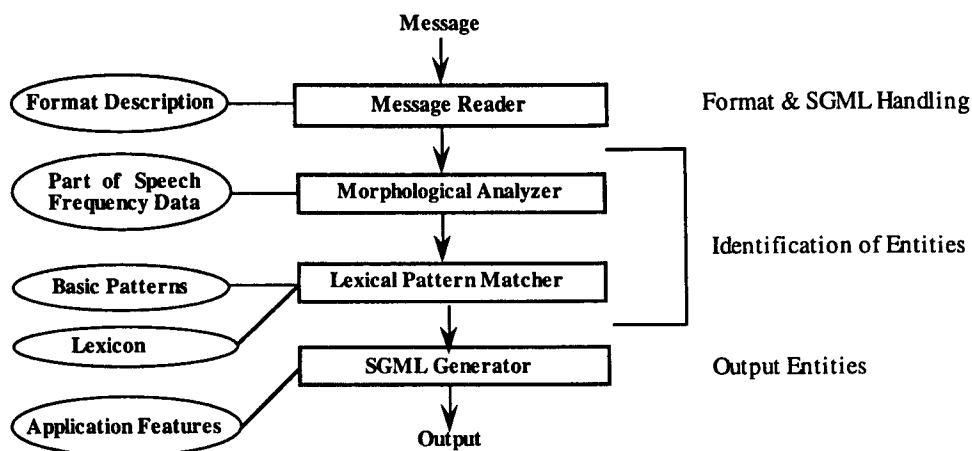

Figure 1: NE System Architecture: *Rectangles represent domain-independent, language-independent algorithms; ovals represent knowledge bases*

The three MUC-6 systems represent three different levels of complexity. The more complex systems are built on top of the simpler systems in order to minimize duplication of effort and maximize knowledge transfer. The NE task is the simplest task and makes use of only lightweight processes, the first three modules of the PLUM system (the message reader, the morphological analyzer, and the lexical pattern matcher).

The TE task takes the entity names found by the NE system, and merges multiple references to the same entity using syntactic and semantic information. The knowledge bases of TE are inherited by ST and do not include domain-specific knowledge. Domain-specific knowledge is localized only in ST.

The NE system is written completely in C and can either be run as a standalone system or as a server which can be queried by the TE and ST systems, which are written in Lisp.

Message Reader

The input to the PLUM system is a file containing one or more messages. The message reader module determines message boundaries, identifies the message header information, and determines paragraph and sentence boundaries. The standalone NE system uses a different message reader than the TE and ST systems.
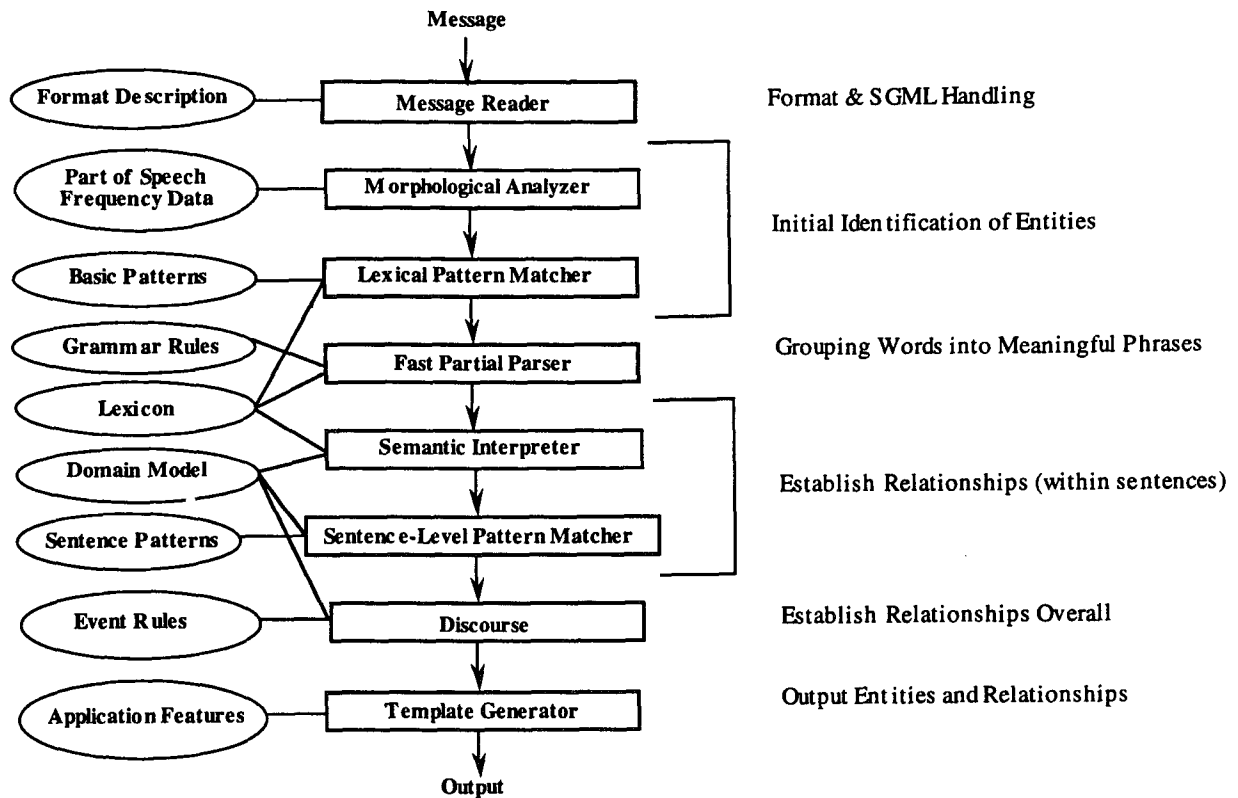
Figure 2: TE/ST System Architecture: *Rectangles represent domain-independent, language-independent algorithms; ovals represent knowledge bases.*

NE

The NE system uses a generic SGML parser to read messages. A parameter file lists the SGML tags relevant to the task, in this case <HL>, <TXT>, <DATELINE>, and <DD>. All other SGML tag pairs are read but ignored.

TE/ST

The TE/ST system uses a more complex message reader. The specification of the input format is declarative, allowing the system to be easily adapted to handle different message formats. This more sophisticated reader can not only pass the portions of the message on to the rest of the system for processing, but can also extract header information (e.g. the document number) from the message and save that information to become part of a template.

Morphological Analyzer

The first phase of processing in both the C and Lisp systems is assignment of part-of-speech information, e.g., proper noun, verb, adjective, etc. In BBN's part-of-speech tagger POST [5], a bi-gram probability model, frequency models for known words (derived from large corpora), and probabilities based on word endings for unknown words are employed to assign part of speech to the highly ambiguous words and unknown words of the corpus. POST tags each word with one of 47 possible tags with 97% accuracy for known words. Below are the part-of-speech tags produced by POST for the following sentence from the walkthrough message:

"And concentrate on his duties as rear commodore at the New York Yacht Club."

((And CC) (concentrate VB) (on IN) (his PP$) (duties NNS) (as IN) (rear JJ) (commodore NN) (at IN) (the DT) (New York NP) (Yacht NP) (Club NP) (. .))

Lexical Pattern Matcher

The Lexical Pattern Matcher was developed after MUC-4 to deal with grammatical forms, such as corporation names. It applies finite state patterns to the input, which consists of word tokens with part-of-speech and semantic

concept information. In particular, word groups that are important to the domain and that may be detectable with only local syntactic analysis can be treated here. When a pattern is matched, a semantic form is assigned by the pattern. In the NE system, patterns were used to recognize all three of the expression types which make up the task (entity expressions, temporal expressions, and numerical expressions). The TE and ST systems gather the results of the NE system's processing and incorporate them in the form of lexicon additions. The TE and ST systems contain no lexical patterns of their own, relying entirely on the domain-independent patterns within the NE system.

Continuing with the example sentence discussed above, a pattern recognized the sequence (New York NP) (Yacht NP) (Club NP) as an organization; the pattern's action substituted the single token (New York Yacht Club NP) with semantics of organization. The Lexical Pattern Matcher is the final step in the processing done by the NE system. The set of recognized entities is used by the output functions to SGML-mark the input.

Fast Partial Parser (FPP)

The FPP is a near-deterministic parser which generates one or more non-overlapping parse fragments spanning the input sentence, deferring any difficult decisions on attachment ambiguities. When cases of permanent, predictable ambiguity arise, the parser finishes the analysis of the current phrase and begins the analysis of a new phrase. Therefore, the entities mentioned and some relations between them are processed in every sentence, whether syntactically ill-formed, complex, novel, or straightforward. Furthermore, this parsing is done using essentially domain-independent syntactic information.

FPP averages about 10 fragments for sentences as complex as in the ST corpus; this number is inflated since punctuation usually results in an isolated fragment. Figure 3 shows parse fragments for two sentences which generated the bulk of the succession information in the walkthrough message.

```
F1: "Mr. James"                              (VP (V "retire")
    (NP (NP (N "Mr.")) (NP (N (NAME "James"))))    (PP (PREP "as")
F2: ","                                              (NP (N "chairman")))))
    (PUNCT ",")                         F10: "at the end of the year"
F3: "57 years"                              (PP (PREP "at")
    (NP (DETERMINER (NUM "57")) (N "years"))    (NP (NP (DETERMINER "the") (N "end"))
F4: "old"                                       (PP (PREP "of")
    (ADJP (ADJ "old"))                              (NP (DETERMINER "the")
F5: ", is stepping down"                             (N "year")))))
    (VP (PUNCT ",")                     F11: "."
        (VP (AUX (V "is"))                  (PUNCT ".")
            (VP (V "stepping down"))))
F6: "as chief executive officer"        F1: "He will be succeeded by Mr. Dooner, 45 ."
    (PP (PREP "as")                         (S (NP (PRO-DET-SPEC "He"))
        (NP (N "chief executive officer")))     (VP (AUX (MODAL "will") (V "be"))
F7: "on July 1"                                 (VP (V "succeeded") (NP)
    (PP (PREP "on")                                 (PP (PREP "by")
        (NP (MONTH "July") (NUM "1")))                  (NP (NP (N "Mr."))
F8: "and"                                                   (NP (NP (N (NAME "Dooner")))
    (CONJ "and")                                            (PUNCT ",")
F9: "will retire as chairman"                               (NUM "45")
    (VP (AUX (MODAL "will"))                                (PUNCT ".")))))))
```

Figure 3. Parser Output: Partial parse found for the example sentences.

Semantic Interpreter

The semantic interpreter contains two sub-components: a rule-based fragment interpreter and a pattern-based sentence interpreter. The first has been used since MUC-3. The rule-based fragment interpreter applies semantic rules to each fragment produced by FPP in a bottom-up, compositional fashion. Semantic rules are matched based on general syntactic patterns, using wildcards and similar mechanisms to provide robustness. A semantic rule creates a semantic representation of the phrase as an annotation on the syntactic parse. A semantic formula includes a variable (e.g., ?13), its type and a collection of predicates on that variable. There are three basic types of semantic forms: entities, events, and states of affairs. Each of these can be further categorized as known, unknown, and referential. Entities correspond to the people, places, things, and time intervals of the domain. These are related in various ways, such as through events (who did what to whom) and states of affairs (properties of the entities). Entity descriptions typically arise from noun phrases; events and states of affairs are often described in clauses.

The rule-based fragment interpreter encodes defaults so that missing semantic information does not produce errors, but marks elements or relationships as unknown. Partial understanding is critical to text processing systems, as missing data is normal. For example, the generic predicate PP-MODIFIER indicates that two entities are connected via a certain preposition. In this way, the system has a "placeholder" for the information that a certain structural relation holds, even though it does not know what the semantic relation is. Sometimes understanding the relation more fully is of no consequence, since the information does not contribute to the template-filling task. The information is maintained, however, so that later expectation-driven processing can use it if necessary.

Due to the fragmentation produced by FPP, top-level constituents are typically more shallow and less varied than full sentence parses. A fairly high level of semantics coverage can be obtained quite quickly when the system is moved to a new domain. This would not be possible if the semantic rules were required to cover a wider variety of syntactic structures before it could achieve reasonable performance. In this way, semantic coverage can be added gradually, while the rest of the system is progressing in parallel.

The second sub-component of the semantic interpreter module is a pattern-based sentence interpreter which applies semantic pattern-action rules to the semantics of each fragment of the sentence. The semantic pattern matching component employs the same core engine as the lexical pattern matcher. These semantic rules can add additional long-distance relations between semantic entities in different fragments within a sentence. A typical TE-level pattern would seek to attach descriptions to organizations, while a ST-level rule would find a potential SUCCESSION and attach the PERSON, ORGANIZATION and POST information related to it. Below is an example of a sentence-level rule which looks for the pattern [<PERSON> ... <JOB-SITUATION-WORD>...<POST>+]. When matched against the sentence "Mr. James, 57 years old, is stepping down as chief executive officer on July 1 and will retire as chairman at the end of the year.", "James" would be <PERSON>, "stepping down" would be <JOB-SITUATION-WORD> and "chief executive officer" would be <POST>.

```
(def-sp-top-patt DEFAULT-PROMOTED
(:pattern
        (:seq
        (:star :anyword)
        (:rule NAMED-PERSON $person)
        (:dontcare 0 2)
        (:tag $succ (:and-env (:cat (VP) :high)
        (:concept JOB-SITUATION)))
        (:dontcare 0 2)
        (:macro (MULTI-POST))
        (:star :anyword)))
(:understanding
        (((:type JOB-SITUATION $succ)
        (:pred JOB-SITUATION-PERSON $succ $person)
        (:pred JOB-SITUATION-POSITION-PERSON $succ $post1)...)))
```

The semantic lexicon is separate from the parser's lexicon and has much less coverage. We use multiple levels of semantic lexicons: first, a generic application-independent lexicon with very shallow semantic information, then the TE lexicon which provides more detailed entity-related semantic information, and finally the ST-level lexicon which provides detailed succession-related entries. Lexical semantic entries indicate the word's semantic type (a domain model concept), as well as predicates pertaining to it. For example, here is the lexical semantic entry for "step down":

```
(DEFVERB "step down"
  (STEP-DOWN-V-1 JOB-SITUATION
                (:preds (JOB-SITUATION-STATUS :self (make-form POSITION-STATUS-GEN-OUT))
                        (JOB-SITUATION-REASON :self (make-form RESIGNATION)))
                (:CASE (LOGICAL-SUBJECT PERSON JOB-SITUATION-PERSON)
                        ("AS" JOB-POSITION JOB-SITUATION-POSITION)
                        ("AS" TITLED-PERSON JOB-SITUATION-POSITION-PERSON)))
  (STEP-DOWN-V-2 JOB-SITUATION
                (:preds (JOB-SITUATION-REASON :self (make-form RESIGNATION)))
                (:CASE (LOGICAL-SUBJECT JOB-POSITION JOB-SITUATION-POSITION)
                        (LOGICAL-SUBJECT TITLED-PERSON JOB-SITUATION-POSITION-PERSON))))
```

In Figure 4, we show the semantic representation that is built for the sentence "He will be succeeded by Mr. Dooner, 45." in the walkthrough article. The whole sentence was parsed as a single fragment by FPP. The JOB-SITUATION event is a flat object combining information from both SUCCESSION and IN-AND-OUT objects for ease of processing. The POSITION-STATUS-GEN-IN state of affairs indicates that the system is unsure of whether the status should be "IN" or "IN_ACTING". Note the link between the "Unknown role" in the JOB-SITUATION and the PERSON "He". The discourse component will resolve the reference for the pronoun and will further refine the relationship between the PERSON and the JOB-SITUATION.
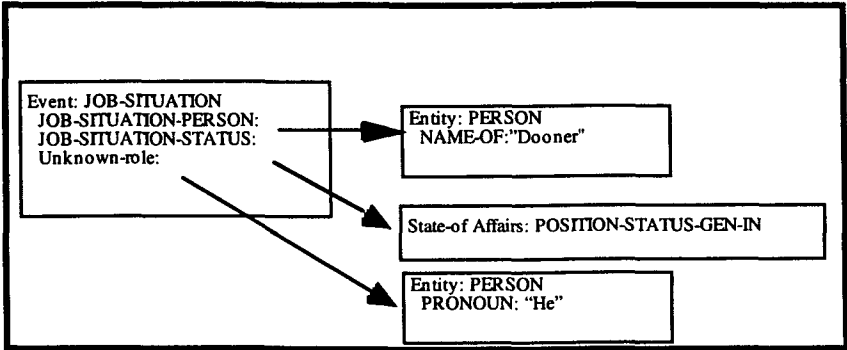

Figure 4. Semantic Structure: *The semantic representation for "He will be succeeded by Mr. Dooner, 45."*

Discourse Processing

PLUM's discourse component creates a meaning for the whole message from the meaning of each sentence. The message level representation is a list of discourse domain objects (DDOs) for the top-level events of interest in the message (e.g., SUCCESSION events in the ST domain). The semantic representation of a phrase in the text only includes information contained nearby; the discourse module must infer other long-distance or indirect relations not explicitly found by the semantic interpreter and resolve any references in the text.

The discourse component creates two primary structures: a discourse predicate database and the DDOs. The database contains all the predicates mentioned in the semantic representation of the message. When references are resolved, corresponding semantic variables are unified. Any other inferences are also added to the database.

To create the DDOs, the discourse component processes each semantic form produced by the interpreter, adding its information to the database. It performs reference resolution for pronouns and anaphoric definite NPs; set- and member-type references may be treated. The discourse component then applies inference rules that may add more semantic information to the discourse predicate database. When a semantic form for an event of interest is encountered, a DDO is generated and any slots already found by the interpreter are filled in. The discourse processor then tries to merge the new DDO with a previous DDO, in order to account for the possibility that the new DDO might be a repeated reference to an earlier one.

Once all the semantic forms have been processed, heuristic rules are applied to fill any empty slots from the text surrounding the forms that triggered a given DDO. Each filler found in the text is assigned a confidence score based on distance from trigger. Fillers found nearby are of high confidence, while those farther away receive worse scores (low numbers represent high confidence; high numbers low confidence; thus 0 is the "highest" confidence score).

In the ST system, the discourse processor performs an additional task. In order to simplify intermediate processing (semantic interpretation, DDO merging, pattern matching), a flat DDO, JOB-SITUATION, was defined which contains the equivalent of the information in a particular SUCCESSION/IN-AND-OUT pair. The flat structure of the JOB-SITUATION object makes merging much simpler as well as making the case frame and semantic patterns easier to define. These must be converted to SUCCESSION and IN-AND-OUT objects before the template generation step, since this is what the template generator expects. The discourse processor performs this conversion.

Each trigger fragment contains one or more words whose semantics triggered the DDO. A DDO can have multiple trigger fragments if the discourse component determines that the triggers co-refer. In this example, "stepping down" in the first fragment and "succeeded" in the second fragment are judged by the discourse processor to be referring to the same succession. A score of 0 indicates that the filler was found either directly by the semantics or by a sentence-level pattern; 1 that it was found in the same fragment as a trigger form; 2 in the same sentence; 4 in the

same paragraph; and 6 in an adjacent paragraph. The "<=>" indicates a reference resolution by the discourse processor. Note that in the first IN-AND OUT, the pronoun "he" has been determined to refer to "James".

A set of DDOs for a SUCCESSION (and its associated IN-AND-OUTs) in the walkthrough message follow:

DDO: SUCCESSION
Trigger fragments:
  ", is stepping down"
  "He will be succeeded by Mr. Dooner, 45 ."

  SUCCESSION-ORG-OF: "McCann-Erickson" (score=4)
  SUCCESSION-POST-OF: "CHIEF EXECUTIVE OFFICER" (score=0)
                        "CHAIRMAN" (score=1)
                        "PRESIDENT" (score=6)
                        "CHIEF OPERATING OFFICER" (score=6)
  SUCCESSION-IN-AND-OUT-OF: IN-AND-OUT-2719 (score=0)
                              IN-AND-OUT-2723 (score=0)
  SUCCESSION-VACANCY-REASON-OF: RESIGNATION (score= 0)
                                  RETIREMENT (score= 2)

DDO: IN-AND-OUT-2719:
  ", is stepping down"

  IN-AND-OUT-PERSON-OF: "James" (score=0) <=>   "he"
                                                         "James"
                                                         "Robert L. James"
                              "McCann" (score= 2)
                              "official" (score= 2)
                              "Dooner" (score= 4) <=>   "John J. Dooner Jr."
                              "One" (score= 6) <=> "Kevin Goldman"
  IN-AND-OUT-NEW-STATUS-OF: POSITION-STATUS-GEN-OUT (score= 0)
                              POSITION-STATUS-GEN-IN (score=4)

DDO:IN-AND-OUT-2723:
  "He will be succeeded by Mr. Dooner, 45 ."

  IN-AND-OUT-PERSON-OF: "Dooner" (score= 0)
                              "McCann" (score= 4)
                              "official" (score= 4)
                              "James" (score= 4) <=>   "Robert L. James"
                              "One" (score= 6) <=>   "Kevin Goldman"
  IN-AND-OUT-NEW-STATUS-OF: POSITION-STATUS-GEN-IN (score= 0)
                              POSITION-STATUS-GEN-OUT (score= 4)

Template Generation

The template generator takes the DDOs produced by discourse processing and fills out the application-specific templates. Clearly, much of this process is governed by the specific requirements of the application, considerations which have little to do with linguistic processing. The template generator must address any arbitrary constraints, as well as deal with the basic details of formatting.

The template generator uses a combination of data-driven and expectation-driven strategies. First the DDOs found by the discourse module are used to produce template objects. Next, the slots in those objects are filled using information in the DDO, the discourse predicate database, other sources of information such as the message header (e.g., document number), or from heuristics (e.g., the type of an organization object is most likely to be COMPANY). The template definitions for the objects which are common to both TE and ST are almost identical. However, they sometimes differ due to the fact that different heuristics or template filling strategies may result in better performance in each of the domains.

61

Parameters in PLUM

Many aspects of PLUM's behavior can be controlled by simply varying the values of system parameters. For example, PLUM has parameters to control aspects of tagging, parsing, pattern matching, event merging and slot filling by discourse, and template filling. An important goal has been to make our system as "parameterizable" as possible, so that the same software can meet different demands for recall, precision, and overgeneration. The final parameter settings for the test were generated by running the systems over all of the data we had and choosing the setting which seemed to maximize the value of the F-measure.

TRAINING TECHNIQUES

Half of the training data was set aside for blind test until the last week of the evaluation; the remaining half was for development. This avoids overfitting to the development data. Since so little data was available, we also created our own training data from Wall Street Journal articles from 1987-1992. This was created by retrieving articles using the University of Massachusetts document retrieval engine INQUERY. This gave us more training data (though presumably of a lesser quality) without violating the integrity of our high-quality blind test set. The blind test set was used to measure our progress at least once a week, with the frequency increasing as the end of the evaluation approached. Figure 5 shows how performance on the ST task improved over time on our blind test set.

Early on, we identified key sentences (those sentences directly responsible for the generation of some important entity such as a SUCCESSION). This allowed us to focus development on those portions of the data which were directly relevant to the task without having to always read through the irrelevant portions.
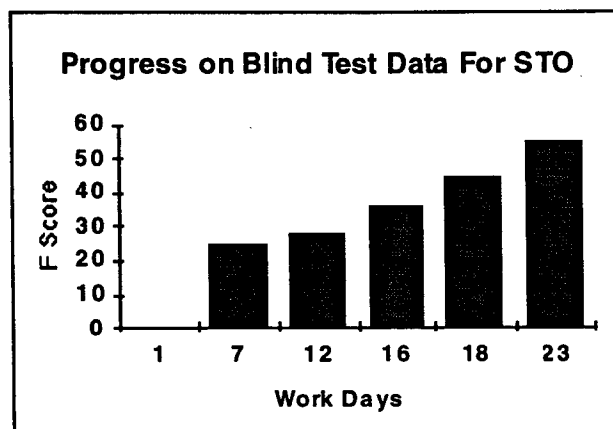

Figure 5: Measured Progress on the New Domain.

We also evaluated system changes on a daily basis using the scores from the training/development set. The scoring program served as a guide for our development. Also, the message-by-message output allowed us to zero in on messages where our performance was particularly bad and allowed us to add lexical items or semantic interpretation rules based on the key sentences in the message. Since the new scoring software did not support visualization of the differences between system output and the answer key, we wrote a visualization tool to do so.

In all evaluation tasks (NE, TE, and ST), PLUM was run over all messages to detect and correct any causes of system breaks. During the last week, we "opened" the blind material which had been released in early September. The entity name slot for all messages was used to quickly add names to the domain-dependent lexicon for TE and ST. The texts were examined for words or collocations which needed to be added to the domain-dependent lexicon. An automated tool generated n-tuples based on part-of-speech tag to aid this process.

TE was slightly different as the training data from the dry run was still valid. Despite this, once the new training data arrived we concentrated almost exclusively on it, mainly using the older data as a "sanity check" before making system changes. We believed that this was important because the nature of the messages from the dry run was quite different than that of the test messages because they had been chosen based on their relevance to the ST domain.

62

	F	Recall	Precision
NE	93.65	94	93
TE	71.97	66	79
ST	54.39	50	59

Figure 6: PLUM System Official Results

A FUNNY THING HAPPENED ON THE WAY TO MUC-6

Early in the planning of MUC-6, an additional dimension for evaluating parsers was planned. We prepared SPATTER for such an evaluation, and had achieved quite high scores on blind test material. However, the parsing evaluation was eventually canceled after the dry run for MUC-6, held in April-May.[1] At that point, we decided that the limited resources we had for MUC should be devoted exclusively to improve scores on the application tasks (NE, TE, and ST), rather than trying to integrate the SPATTER parser for the application evaluations.

We look forward to integrating SPATTER in future information extraction tasks to test the hypothesis that a far more accurate parser could lead to more accurate understanding and to notably higher scores.

LIMITING FACTORS

Three factors significantly limited us. The first was the relative lack of data provided. Providing twice the messages marked for NE, TE, and ST would have made a big difference. Consequently, we created our own additional data and answer keys for NE and ST. Given so few messages, we felt that there were conventions in filling out the keys in each task that were still not fully clear. In fact, with students experienced in marking NE, consistency across annotators was only 94%, suggesting that the annotation rules can use further elaboration.

Due to ongoing application efforts with tight deadlines, the limited availability of experienced MUCcaneers, and prior investment in software to find names, we put in less effort than on any of the MUCs (-3, -4, and -5) which we had previously participated in.

As a consequence of needing to limit the effort that we could give, we decided to focus on ST more than the other two tasks. TE received the least effort.

LESSONS LEARNED (CONCLUSIONS)

Regarding NE

Several conclusions seem warranted from NE.

1. The most exciting is that **near human performance is within the state of the art for mixed case English**. Several systems performed at 90% or above.
2. Our next steps are to improve IdentiFinder's prediction of aliases once a name has been seen and to add rules for low frequency cases, e.g., improving performance on names that are quite unlike Western European names.
3. **We would encourage looking at harder cases for NE evaluation.** In Wall Street Journal, NE is substantially simplified by accurate usage of mixed case. How would these systems perform in upper case only or in languages where initial capitalization does not signal a name? Languages such as Japanese and Chinese have no capital letters; languages such as German use capitalization for all nouns, not just nouns in names.

Regarding TE

We conclude the following:

1. As with NE, **many groups performed at a level higher than any previous template fill task in MUC-3, -4, or -5.** It will be interesting to see if this general template task is broadly useful, and whether performance is at a level high enough to warrant deployment in some real task(s).

---

[1] Given the already ambitious nature of MUC-6, we do not disagree with the decision to consolidate on the four evaluation tasks, nor are we arguing to make parsing an evalution task in MUC-7.

2. Given the little effort we invested in TE, we believe that another two person weeks could get PLUM's scores on blind test material over 80. To do this, we would improve NE performance (see discussion of NE above) and would work further on the locale, country, and descriptor slots. An example of simple alias improvements is to recognize "Mr. Smith" as an alias after seeing "John Smith," instead of merely predicting "Smith" and "John" as possible aliases. We also believe that it is possible to achieve an F of 80 or better in TE using only lightweight processing.

3. To achieve an F above 90, on the other hand, is likely to require significant overall improvement and heavyweight processing in particular. We would incorporate the SPATTER parser, which parses far more accurately than FPP does, and would look to add to our domain-independent semantic lexicon, so that there is more semantic information to support merging of entity descriptions. Together, we believe these offer the best prospect for radically improved performance in the descriptor, locale and country slots.

4. Our TE system by design, employs no domain-specific knowledge. **It therefore should work equally well on other text, not specific to change in corporate officers.**

5. **PLUM's Performance had not peaked on TE**, since we put the least effort on it of any of the evaluation dimension. Furthermore, the official test was not even reflective of PLUM's performance, since a set of rule variations that was known to improve performance was saved at the ST level, rather than at the TE level (which ST would have inherited). The effect of our mistake and others was substantial and is given in the table below.

	F	Recall	Precision
Official	71.97	66	79
Unofficial	76.46	72	81

Regarding ST

Several conclusions seem warranted from ST.

1. **Compared to PLUM's previous performance in MUC-3, -4, and -5, our progress was much more rapid and our official score was higher than in any previous template fill task.**

2. The division of objects into broadly applicable ones (TE) and domain-specific ones (ST) was a plus in our opinion. It felt as though the amount of domain-specific overhead was lowered, compared to previous MUCs.

3. Given that, we urge an even simpler template structure for future MUCs, one where only two levels are present: entities and relationships. This would more closely match what is stated in the text and would factor out the issues of data base structure. Together that should reduce overhead for participants further.

4. Though we believe that an additional 5-10 point improvement in F would have been achievable with more calendar time than 30 days for the ST task, to achieve an F above 65 is likely to require significant overall improvement. We would like to incorporate the SPATTER parser, which parses far more accurately than FPP does, and would look to add to our domain-independent semantic lexicon so as to improve merging of entity descriptions. Together, this should improve overall recognition, merging and discrimination of all objects, and may be a key to accurately recognizing events/relationships not detectable in a single sentence.

OVERALL CONCLUSIONS

We make two general observations. First, the state of the art has progressed greatly in portability in the last four years. For MUC-3, some high performing groups invested a small number of person years. By contrast, several groups this year achieved an F in the 50s in 30 calendar days.

Yet, we believe that we are only beginning to understand techniques for learning domain-independent knowledge and domain-dependent knowledge. Far more can be achieved. BBN particularly would like to investigate how statistical algorithms over large unmarked corpora can effectively extrapolate from a few training examples, such as in ST in MUC-6, to provide greater coverage. For example, statistical techniques may have suggested the importance of "hire," a verb which many groups did not define.

Second, since there has been a marked improvement in the quality of full parsers, now achieving an F in the high 80s (Magerman, 1995), we believe it is now feasible to consider using full parsers again. The rationale is straightforward: for full templates (e.g., ST) scores have been mired with an F in the 50s ever since MUC-3 in 1991. Pattern matching has given us very robust, very portable technology, but has not broken the performance barrier all systems have run up against. Recent (statistical) full parsers (e.g., BBN's, IBM's, and UPenn's) have such quantitatively better performance that they are qualitatively better. We believe this offers the opportunity to again

try heavyweight techniques to attempt deeper understanding. Pattern matching techniques will still have a crucial role for domain-specific details, but we believe they can be greatly improved by deeper understanding.

SYSTEM WALKTHROUGHS

No development was done on the walkthrough messages for any of the domains. The walkthrough message has 3 succession events. The first 2 successions come from 2 sentences (s1 & s2), where Mr. Dooner replaces Mr. James in two positions (CEO & chairman). The third succession comes from a much later sentence (s3) having to do with a vice-presidency. (s1) alone does not completely define one of the successions (each of which has 2 in-and-out objects for the 2 people involved).

"Now, Mr. James is preparing to sail into the sunset, and Mr. Dooner is poised to rev up the engines to guide Interpublic Group's McCann-Erickson into the 21st century. Yesterday, McCann made official what had been widely anticipated:
(s1) Mr. James, 57 years old, is stepping down as chief executive officer on July 1 and will retire as chairman at the end of the year.
(s2) He will b; succeeded by Mr. Dooner, 45."

PLUM performed fairly well on this article. We found the right corporations, the right people, the right positions, etc. When we saw "person p1 retires from positions y and z" and "person p2 will succeed him" we don't infer p2 takes over both positions, and in this case, we pair his succeeding with only position y, thus the first succession is completely correct and gets the 2 in-and-out objects, but the second only has one. In the TE output, we missed the alias "John Dooner", possibly due to a shortcoming in the aliasing algorithm.

NE OUTPUT: (for s1, s2)

Mr. <ENAMEX TYPE="PERSON">James</ENAMEX>, 57 years old, is stepping down as chief executive officer on <TIMEX TYPE="DATE">July 1</TIMEX> and will retire as chairman at the end of the year. He will be succeeded by Mr. <ENAMEX TYPE="PERSON">Dooner</ENAMEX>, 45.

TE OUTPUT: (for s1, s2)

<PERSON-9402240133-2> :=
   PER_NAME: "Robert L. James"
   PER_TITLE: "Mr."
   PER_ALIAS: "James"

<PERSON-9402240133-3> :=
   PER_NAME: "John J. Dooner Jr."
   PER_TITLE: "Mr."
   PER_ALIAS: "Dooner"

ST OUTPUT: (for s1, s2)

<SUCCESSION_EVENT-9402240133-1> :=
   SUCCESSION_ORG: <ORGANIZATION-9402240133-1>
   POST: "CHIEF EXECUTIVE OFFICER"
   IN_AND_OUT: <IN_AND_OUT-9402240133-1>
              <IN_AND_OUT-9402240133-3>
   VACANCY_REASON: REASSIGNMENT
<SUCCESSION_EVENT-9402240133-2> :=
   SUCCESSION_ORG: <ORGANIZATION-9402240133-1>
   POST: "CHAIRMAN"
   IN_AND_OUT: <IN_AND_OUT-9402240133-2>
   VACANCY_REASON: DEPART_WORKFORCE
<IN_AND_OUT-9402240133-1> :=
   IO_PERSON: <PERSON-9402240133-1>
   NEW_STATUS: OUT
   ON_THE_JOB: UNCLEAR
<IN_AND_OUT-9402240133-2> :=
   IO_PERSON: <PERSON-9402240133-1>

NEW_STATUS: OUT
ON_THE_JOB: UNCLEAR
<IN_AND_OUT-9402240133-3> :=
   IO_PERSON: <PERSON-9402240133-2>
   NEW_STATUS: IN
   ON_THE_JOB: UNCLEAR
<ORGANIZATION-9402240133-1> :=
   ORG_NAME: "McCann-Erickson"
   ORG_TYPE: COMPANY
<PERSON-9402240133-1> :=
   PER_NAME: "Robert L. James"
   PER_TITLE: "Mr."
   PER_ALIAS: "James"
<PERSON-9402240133-2> :=
   PER_NAME: "John J. Dooner Jr."
   PER_TITLE: "Mr."
   PER_ALIAS: "Dooner"

FULL NE OUTPUT:

<DOC>
<DOCID> wsj94_026.0231 </DOCID>
<DOCNO> 940224-0133. </DOCNO>
<HL> Marketing & Media—Advertising:

@ John <ENAMEX TYPE="PERSON">Dooner</ENAMEX> Will Succeed <ENAMEX TYPE="PERSON">James</ENAMEX>
@ At Helm of <ENAMEX TYPE="ORGANIZATION">McCann-Erickson</ENAMEX>
@ ----
@ By <ENAMEX TYPE="PERSON">Kevin Goldman</ENAMEX> </HL>
<DD> <TIMEX TYPE="DATE">02/24/94</TIMEX> </DD>
<SO> WALL STREET JOURNAL (J), PAGE B8 </SO>
<CO>   IPG K </CO>
<IN> ADVERTISING (ADV), ALL ENTERTAINMENT & LEISURE (ENT), FOOD PRODUCTS (FOD), FOOD PRODUCERS, EXCLUDING FISHING (OFP), RECREATIONAL PRODUCTS & SERVICES (REC), TOYS (TMF) </IN>
<TXT>
<p>
One of the many differences between <ENAMEX TYPE="PERSON">Robert L. James</ENAMEX>, chairman and chief executive officer of <ENAMEX TYPE="ORGANIZATION">McCann-Erickson</ENAMEX>, and <ENAMEX TYPE="PERSON">John J. Dooner Jr.</ENAMEX>, the agency's president and chief operating officer, is quite telling: Mr. <ENAMEX TYPE="PERSON">James</ENAMEX> enjoys sailboating, while Mr. <ENAMEX TYPE="PERSON">Dooner</ENAMEX> owns a powerboat.
</p>
<p>
Now, Mr. <ENAMEX TYPE="PERSON">James</ENAMEX> is preparing to sail into the sunset, and Mr. <ENAMEX TYPE="PERSON">Dooner</ENAMEX> is poised to rev up the engines to guide <ENAMEX TYPE="ORGANIZATION">Interpublic Group</ENAMEX>'s <ENAMEX TYPE="ORGANIZATION">McCann-Erickson</ENAMEX> into the 21st century. Yesterday, <ENAMEX TYPE="ORGANIZATION">McCann</ENAMEX> made official what had been widely anticipated: Mr. <ENAMEX TYPE="PERSON">James</ENAMEX>, 57 years old, is stepping down as chief executive officer on <TIMEX TYPE="DATE">July 1</TIMEX> and will retire as chairman at the end of the year. He will be succeeded by Mr. <ENAMEX TYPE="PERSON">Dooner</ENAMEX>, 45.
</p>
<p>
It promises to be a smooth process, which is unusual given the volatile atmosphere of the advertising business. But Mr. <ENAMEX TYPE="PERSON">Dooner</ENAMEX> has a big challenge that will be his top priority. "I'm going to focus on strengthening the creative work," he says. "There is room to grow. We can make further improvements in terms of the perception of our creative work."
</p>
<p>
Even <ENAMEX TYPE="PERSON">Alan Gottesman</ENAMEX>, an analyst with <ENAMEX TYPE="ORGANIZATION">PaineWebber</ENAMEX>, who believes <ENAMEX TYPE="ORGANIZATION">McCann</ENAMEX> is filled with "vitality" and is in "great shape," says that from a creative standpoint, "You wouldn't pay to see their reel" of commercials.
</p>
<p>
While <ENAMEX TYPE="ORGANIZATION">McCann</ENAMEX>'s world-wide billings rose <NUMEX TYPE="PERCENT">12%</NUMEX> to <NUMEX TYPE="MONEY">$6.4 billion</NUMEX> last year from <NUMEX TYPE="MONEY">$5.7 billion</NUMEX> in <TIMEX TYPE="DATE">1992</TIMEX>, the agency still is dogged by the loss of the key creative assignment for the prestigious <ENAMEX TYPE="ORGANIZATION">Coca-Cola</ENAMEX> Classic account. "I would be less than honest to say I'm not disappointed not to be able to claim creative leadership for <ENAMEX TYPE="ORGANIZATION">Coke</ENAMEX>," Mr. <ENAMEX TYPE="PERSON">Dooner</ENAMEX> says.
</p>
<p>
<ENAMEX TYPE="ORGANIZATION">McCann</ENAMEX> still handles promotions and media buying for <ENAMEX TYPE="ORGANIZATION">Coke</ENAMEX>. But the bragging rights to <ENAMEX TYPE="ORGANIZATION">Coke</ENAMEX>'s ubiquitous advertising belongs to <ENAMEX TYPE="ORGANIZATION">Creative Artists Agency</ENAMEX>, the big <ENAMEX TYPE="LOCATION">Hollywood</ENAMEX> talent agency. "We are striving to have a strong renewed creative partnership with <ENAMEX TYPE="ORGANIZATION">Coca-Cola</ENAMEX>," Mr. <ENAMEX TYPE="PERSON">Dooner</ENAMEX> says. However, odds of that happening are slim since word from

<ENAMEX TYPE="ORGANIZATION">Coke</ENAMEX> headquarters in <ENAMEX TYPE="LOCATION">Atlanta</ENAMEX> is that CAA and other ad agencies, such as <ENAMEX TYPE="PERSON">Fallon McElligott</ENAMEX>, will continue to handle <ENAMEX TYPE="ORGANIZATION">Coke</ENAMEX> advertising.
</p>
<p>
Mr. <ENAMEX TYPE="PERSON">Dooner</ENAMEX>, who recently lost <NUMEX TYPE="MONEY">60 pounds</NUMEX> over three-and-a-half months, says now that he has "reinvented" himself, he wants to do the same for the agency. For Mr. <ENAMEX TYPE="PERSON">Dooner</ENAMEX>, it means maintaining his running and exercise schedule, and for the agency, it means developing more global campaigns that nonetheless reflect local cultures. One <ENAMEX TYPE="ORGANIZATION">McCann</ENAMEX> account, "I Can't Believe It's Not Butter," a butter substitute, is in 11 countries, for example.
</p>
<p>
<ENAMEX TYPE="ORGANIZATION">McCann</ENAMEX> has initiated a new so-called global collaborative system, composed of world-wide account directors paired with creative partners. In addition, <ENAMEX TYPE="PERSON">Peter Kim</ENAMEX> was hired from <EN/ MEX TYPE="ORGANIZATION">WPP Group</ENAMEX>'s <ENAMEX TYPE="PERSON">J. Walter Thompson</ENAMEX> last <TIMEX TYPE="DATE">September</TIMEX> as vice chairman, chief strategy officer, world-wide.
</p>
<p>
Mr. <ENAMEX TYPE="PERSON">Dooner</ENAMEX> doesn't see a creative malaise permeating the agency. He points to several campaigns with pride, including the Taster's Choice commercials that are like a running soap opera. "It's a <NUMEX TYPE="MONEY">$19 million</NUMEX> campaign with the recognition of a <NUMEX TYPE="MONEY">$200 million</NUMEX> campaign," he says of the commercials that feature a couple that must hold a record for the length of time dating before kissing.
</p>
<p>
Even so, Mr. <ENAMEX TYPE="PERSON">Dooner</ENAMEX> is on the prowl for more creative talent and is interested in acquiring a hot agency. He says he would like to finalize an acquisition "yesterday. I'm not known for patience."
</p>
<p>
Mr. <ENAMEX TYPE="PERSON">Dooner</ENAMEX> met with <ENAMEX TYPE="PERSON">Martin Puris</ENAMEX>, president and chief executive officer of <ENAMEX TYPE="ORGANIZATION">Ammirati</ENAMEX> & Puris, about <ENAMEX TYPE="ORGANIZATION">McCann</ENAMEX>'s acquiring the agency with billings of <NUMEX TYPE="MONEY">$400 million</NUMEX>, but nothing has materialized. "There is no question," says Mr. <ENAMEX TYPE="PERSON">Dooner</ENAMEX>, "that we are looking for quality acquisitions and <ENAMEX TYPE="ORGANIZATION">Ammirati</ENAMEX> & <ENAMEX TYPE="PERSON">Puris</ENAMEX> is a quality operation. There are some people and entire agencies that I would love to see be part of the <ENAMEX TYPE="PERSON">McCann</ENAMEX> family." Mr. <ENAMEX TYPE="PERSON">Dooner</ENAMEX> declines to identify possible acquisitions.
</p>
<p>
Mr. <ENAMEX TYPE="PERSON">Dooner</ENAMEX> is just gearing up for the headaches of running one of the largest world-wide agencies. (There are no immediate plans to replace Mr. <ENAMEX TYPE="PERSON">Dooner</ENAMEX> as president; Mr. <ENAMEX TYPE="PERSON">James</ENAMEX> operated as chairman, chief executive officer and president for a period of time.) Mr. <ENAMEX TYPE="PERSON">James</ENAMEX> is filled with thoughts of enjoying his three hobbies: sailing, skiing and hunting.
</p>
<p>
Asked why he would choose to voluntarily exit while he still is so young, Mr. <ENAMEX TYPE="PERSON">James</ENAMEX> says it is time to be a tad selfish about how he spends his days. Mr. <ENAMEX TYPE="PERSON">James</ENAMEX>, who has a reputation as an extraordinarily tough taskmaster, says that because he "had a great time" in advertising," he doesn't want to "talk about the disappointments." In fact, when he is asked his opinion of the new batch of <ENAMEX TYPE="ORGANIZATION">Coke</ENAMEX> ads

from CAA, Mr. <ENAMEX TYPE="PERSON">James</ENAMEX> places his hands over his mouth. He shrugs. He doesn't utter a word. He has, he says, fond memories of working with <ENAMEX TYPE="ORGANIZATION">Coke</ENAMEX> executives. "<ENAMEX TYPE="ORGANIZATION">Coke</ENAMEX> has given us great highs," says Mr. <ENAMEX TYPE="PERSON">James</ENAMEX>, sitting in his plush office, filled with photographs of sailing as well as huge models of, among other things, a Dutch tugboat.
</p>
<p>
He says he feels a "great sense of accomplishment." In 36 countries, <ENAMEX TYPE="PERSON">McCann</ENAMEX> is ranked in the top three; in 75 countries, it is in the top 10.
</p>
<p>
Soon, Mr. <ENAMEX TYPE="PERSON">James</ENAMEX> will be able to compete in as many sailing races as he chooses. And concentrate on his duties as rear commodore at the <ENAMEX TYPE="ORGANIZATION">New York Yacht Club</ENAMEX>.
</p>
<p>
Maybe he'll even leave something from his office for Mr. <ENAMEX TYPE="PERSON">Dooner</ENAMEX>. Perhaps a framed page from the <ENAMEX TYPE="ORGANIZATION">New York Times</ENAMEX>, dated <TIMEX TYPE="DATE">Dec. 8, 1987</TIMEX>, showing a year-end chart of the stock market crash earlier that year. Mr. <ENAMEX TYPE="PERSON">James</ENAMEX> says he framed it and kept it by his desk as a "personal reminder. It can all be gone like that."
</p>
</TXT>
</DOC>

FULL TE OUTPUT:

<ORGANIZATION-9402240133-1> :=
  ORG_NAME: "McCann-Erickson"
  ORG_TYPE: COMPANY
<ORGANIZATION-9402240133-2> :=
  ORG_NAME: "New York Yacht Club"
  ORG_TYPE: OTHER
  ORG_LOCALE: Hollywood UNK
  ORG_COUNTRY: United States
<ORGANIZATION-9402240133-3> :=
  ORG_NAME: "PaineWebber"
  ORG_TYPE: COMPANY
<ORGANIZATION-9402240133-4> :=
  ORG_NAME: "Coke"
  ORG_TYPE: COMPANY
  ORG_LOCALE: Atlanta CITY
  ORG_COUNTRY: United States
<ORGANIZATION-9402240133-5> :=
  ORG_NAME: "Creative Artists Agency"
  ORG_ALIAS: "CAA"
  ORG_TYPE: GOVERNMENT
  ORG_LOCALE: Hollywood UNK
  ORG_COUNTRY: United States
<ORGANIZATION-9402240133-6> :=
  ORG_NAME: "Coca-Cola Classic"
  ORG_ALIAS: "Coca-Cola"
  ORG_TYPE: COMPANY

<ORGANIZATION-9402240133-7> :=
  ORG_NAME: "Fallon McElligott"
  ORG_TYPE: COMPANY
  ORG_LOCALE: Atlanta CITY
  ORG_COUNTRY: United States
<ORGANIZATION-9402240133-8> :=
  ORG_NAME: "WPP Group"
  ORG_TYPE: COMPANY
<PERSON-9402240133-1> :=
  PER_NAME: "Kevin Goldman"
<PERSON-9402240133-2> :=
  PER_NAME: "Robert L. James"
  PER_TITLE: "Mr."
  PER_ALIAS: "James"
<PERSON-9402240133-3> :=
  PER_NAME: "John J. Dooner Jr."
  PER_TITLE: "Mr."
  PER_ALIAS: "Dooner"
<PERSON-9402240133-4> :=
  PER_NAME: "McCann"
<PERSON-9402240133-5> :=
  PER_NAME: "Alan Gottesman"
<PERSON-9402240133-6> :=
  PER_NAME: "Peter Kim"
<PERSON-9402240133-7> :=
  PER_NAME: "Martin Puris"

FULL ST OUTPUT

<TEMPLATE-9402240133-1> :=
  DOC_NR: 9402240133
  CONTENT: <SUCCESSION_EVENT-9402240133-1>
      <SUCCESSION_EVENT-9402240133-2>
      <SUCCESSION_EVENT-9402240133-3>

<SUCCESSION_EVENT-9402240133-1> :=
  SUCCESSION_ORG: <ORGANIZATION-9402240133-1>
  POST: "CHIEF EXECUTIVE OFFICER"
  IN_AND_OUT: <IN_AND_OUT-9402240133-1>
      <IN_AND_OUT-9402240133-3>

68

```
VACANCY_REASON: REASSIGNMENT                          ON_THE JOB  UNCLEAR
<SUCCESSION_EVENT-9402240133-2> :=                    <IN_AND OUT 9402240133-4> :=
    SUCCESSION_ORG: <ORGANIZATION-9402240133-1>          IO_PERSON: <PERSON-9402240133-3>
    POST: "CHAIRMAN"                                     NEW_STATUS: IN
    IN_AND_OUT: <IN_AND_OUT-9402240133-2>               ON_THE_JOB: UNCLEAR
    VACANCY_REASON: DEPART_WORKFORCE                 <ORGANIZATION-9402240133-1> :=
<SUCCESSION_EVENT-9402240133-3> :=                      ORG_NAME: "McCann-Erickson"
    SUCCESSION_ORG: <ORGANIZATION-9402240133-2>         ORG_TYPE: COMPANY
    POST: "VICE CHAIRMAN"                            <ORGANIZATION-9402240133-2> :=
    IN_AND_OUT: <IN_AND_OUT-9402240133-4>               ORG_NAME: "WPP Group"
    VACANCY_REASON: OTH_UNK                             ORG_TYPE: COMPANY
<IN_AND_OUT-9402240133-1> :=                          <PERSON-9402240133-1> :=
    IO_PERSON: <PERSON-9402240133-1>                    PER_NAME: "Robert L. James"
    NEW_STATUS: OUT                                     PER_TITLE: "Mr."
    ON_THE_JOB: UNCLEAR                                 PER_ALIAS: "James"
<IN_AND_OUT-9402240133-2> :=                          <PERSON-9402240133-2> :=
    IO_PERSON: <PERSON-9402240133-1>                    PER_NAME: "John J. Dooner Jr."
    NEW_STATUS: OUT                                     PER_TITLE: "Mr."
    ON_THE_JOB: UNCLEAR                                 PER_ALIAS: "Dooner"
<IN_AND_OUT-9402240133-3> :=                          <PERSON-9402240133-3> :=
    IO_PERSON: <PERSON-9402240133-2>                    PER_NAME: "Peter Kim"
    NEW_STATUS: IN
```

ACKNOWLEDGMENTS

REFERENCES

Ayuso, D.M., Boisen, S., Fox, H., Ingria, R., and Weischedel, R. "BBN: Description of the PLUM System as Used for MUC-4", *MUC-4 Proceedings*, 1992.

Iwanska, et.al., "Computational Aspects of Discourse in the Context of MUC-3", *Proceedings of the Third Message Understanding Conference (MUC-3)*, 1991.

Matsukawa, T., Hypothesizing Word Association from Untagged Text", *Proceedings of the ARPA Workshop on Human Language Technology*, 1993.

Matsukawa, T., Miller, S., and Weischedel, R. "Example-Based Correction of Word Segmentation and Part of Speech Labelling", *Proceedings of the ARPA Workshop on Human Language Technology*, 1993.

Weischedel, R., Ayuso, D.M., Boisen, S., Fox, H., Ingria, R., and Palmucci, J., "Partial Parsing, A Report on Work in Progress", *Proceedings of the Fourth ARPA Workshop on Speech and Natural Language*, 1991.

Weischedel, R., Ayuso, D.M., Bobrow, R., Boisen, S., Fox, H., Matsukawa, T., MacLaughlin, D., Papageorgiou, C., Sakai, T., Abe, J., Hoshi, H., Miyamoto, Y., and Miller, S., "BBN's PLUM Probabilistic Language Understanding System", *Proceedings of the TIPSTER Text Program (Phase I)*, 1993.

Weischedel, R., Meteer, M., Schwartz, R., Ramshaw, L., and Palmucci, J. "Coping with Ambiguity and Unknown Words through Probabilistic Models", *Computational Linguistics (Special Issue on Using Large Corpora: II)* 19, 359-382, 1993.