LOCKHEED MARTIN: LOUELLA PARSING, AN NLTOOLSET SYSTEM FOR MUC-6

by
Lois Childs, Deb Brady, Louise Guthrie, Jose Franco, Dan Valdes-Dapena, Bill Reid, John Kielty, Glenn Dierkes and Ira Sider

Lockheed Martin
Building 10, Room 1527
P.O. Box 8048
Philadelphia Pa. 19101

BACKGROUND

During the 1980s, General Electric Corporate Research and Development began the design and implementation of a set of text-processing tools known as the NLToolset. This suite of tools developed, over time, in cooperation with a subgroup of the Management and Data Systems Operations component of General Electric Aerospace. Through corporate mergers, this subgroup has become the Language Exploitation Technologies group of the Lockheed Martin Management and Data Systems division. Over the years, the toolset has evolved into a robust set of aids for text analysis. It has been used to build a variety of applications, and the knowledge gained from each application has been utilized to improve the toolset. The LOUELLA PARSING system was designed with the latest version of the NLToolset.

The Lockheed Martin group's LOUELLA PARSING system participated in three of the four MUC-6 evaluations:

1) The evaluation of systems that mark named entities (NE) in a text. In particular, systems were required to mark locations, dates, times, organizations, people, and currency phrases in a text.
2) The evaluation of systems that extract template elements (TE) corresponding to people and organizations from the text. These templates can be thought of as index cards containing information about people and organizations mentioned in a text.
3) The evaluation of systems that extract information about corporate management successions from a text. This task is called the "scenario template" (ST), and requires a deeper anlysis of the text than the other tasks.

This paper will give an overview of our systems, describe our performance on each of the tasks and the walkthrough article, and discuss areas where our systems need to be improved.

LOUELLA's NE system was developed using the "spotters" from the NLToolset. The TE and ST systems were developed using a range of tools for reference resolution, for information extraction, for simple discourse processing, and for template generation, as well as a variety of spotters for spotting entities. Below we describe the processing stages that are used in LOUELLA's NE, TE and ST systems. The first three stages are used in the NE task.

PROCESSING STAGES

The NLToolset contains a core knowledge base, a large sense-disambiguated lexicon, and a variety of text-processing tools for extracting information, organizing information, and generating output. Typically, these tools are sequentially applied to text.

Text Tokenization and Segmentation

LOUELLA uses sequential processing which simplifies the text with each phase. The first text-processing module used is NLlex, a lexical-analyzer-development package which handles the character string to word translation and tokenizes the text. Next, the text segmenter interprets the SGML markers and common punctuation, and stores the text in a structure that holds the original version of the text as a whole; it also stores each section of the text, such as paragraph, sentence, headline, dateline, etc. Throughout processing, the structure holds an original and a "latest" version of each sentence; the "latest" version is updated with each processing phase.

Lexical Look-up

At this point, the text is stored in tokenized form, and any unambiguous names and phrases that are stored in the lexicon are identified. Named entities, such as organizations and people, are stored on an active token list to allow the system to link occurrences of the same entity, based on name variations. Next, each word is analyzed morphologically and tagged with its possible parts of speech, which are found in the lexicon. The following is an example of the first check against the lexicon. Here the phrasal "less than" and the company "Coke" are identified and marked as multi-tokens (MT):

"I would be **[MT{LESS-THAN}: less than]** honest to say I'm not disappointed not to be able to claim creative leadership for **[MT{COCA-COLA}: Coke]**," Mr. Dooner says.

Text Reduction

LOUELLA uses a non-deterministic, lexico-semantic, finite-state pattern matcher. By this definition, we refer to the pattern matcher as a finite-state machine which matches against both syntactic and semantic features of text. The pattern matcher uses a knowledge base of pattern-action rules, grouped in rule packages. These rule packages are applied to the text in successive passes to mark primitive text elements such as time, money, locations, person, and company names. These marked phrases are then reduced into single tokens.

LOUELLA uses eight rule-packages to reduce elements ranging from the most primitive time expressions and organization noun phrases to the more complex IN_AND_OUT objects. This phase of text processing extends from the Named Entity system through to the Template Element, and includes the Scenario Template phase. Here are several examples:

Matches: (TIME-RULE-0 TIME-RULE-36 TIME-RULE-38)
[TIME-ABSOLUTE{36}: Yesterday], McCann made official what had been widely anticipated: Mr. James, 57 years old, is stepping down as chief executive officer on **[DATE{4}: July 1]** and will retire as chairman **[TIME-RELATIVE{38}: at the end of the year]** .

Matches: (PNAME-RULE-2)
Yesterday, McCann made official what had been widely anticipated: **[PNAME{2}: 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 .

Matches: (OTHERNP-RULE-3)
[PPNOUN{3}: He] will be succeeded by Dooner, 45.

Matches: (ORGNP-RULE-3)
But the bragging rights to Coke's ubiquitous advertising belongs to Creative Artists Agency, **[ORGNP{3}: the big Hollywood talent agency]** .

Matches: (ORGANIZATION-RULE-7 PERSON-RULE-0)
Now, **[PERSON{0}: James]** is preparing to sail into the sunset, and **[PERSON{0}: Dooner]** is poised to rev up the engines to guide **[ORGANIZATION{7}: Interpublic Group]** 's **[ORGANIZATION{7}: McCann-Erickson]** into the 21st century.

Matches: (IN_AND_OUT-RULE-0 IN_AND_OUT-RULE-1)
Yesterday, McCann made official what had been widely anticipated: **[IN_AND_OUT{0}: James]** , 57

years old, is stepping down as **[OFFICERTOK{1}: chief executive officer]** on July 1 and will retire as **[OFFICERTOK{1}: chairman]** at the end of the year.

Reference Resolution

Reference resolution is ongoing throughout processing. As soon as a named entity is recognized, it is stored—along with its variations—on an active token list so that variations of the name can be recognized and linked to the original occurrence. When organization names are recognized, they can often be directly linked to their appositives or prenominal phrases. In addition, noun-phrase recognition prompts a backward search through a stack of named entities in order to identify its referent. This search uses several tactics to find the correct referent. If the noun-phrase is semantically rich, a content filter is constructed and compared against content filters for known, named entities. If this is not successful, various heuristics are used based on entity type and position in text.

Information Extraction

LOUELLA uses the same pattern matcher for information extraction that it uses for text reduction; however, there is a difference in the way the pattern matcher is used. While extracting, the pattern matcher is allowed to overlap patterns because it is not changing the text found; it is merely extracting information of interest and sending it to the text organizer. The text organizer tries to assemble the extracted information into a lucid account of events. It performs this assembly by using a model of the domain as delineated in the task specification. For example, it is permissible to have more than one IN_AND_OUT object participating in a SUCCESSION_EVENT, but there must be one SUCCESSION_ORG involved.

Postprocessing

Postprocessing is the final review of the extracted information before the templates are generated. It is LOUELLA's chance to apply any heuristics which may seem helpful to an accurate reporting of information. This part of the system is entirely dependent on the domain and can be customized at will by the developer.

Template Generating

LOUELLA has a template generator which uses an object-oriented mapping script for generating the final template. The script is based on the task specification and contains the path which the template generator should follow through the objects. The script also contains pointers to the functions which print each slot fill.

SYSTEM MODULES

For MUC-6, LOUELLA is comprised of three system modules, one for each of the MUC-6 tasks addressed: Named Entity (NE), Template Element (TE), and Scenario Template (ST). Below, we give a brief description of each of these systems.

The NE System

LOUELLA's Named Entity system is a multi-pass process which builds upon entities which are found in previous passes. In addition to the segmentation and lexical look-up stages of our system, early passes of the reduction phase identify time, date, money, and percent components. The system then searches for locations, knowing that the entities found previously will not be part of the location phrase. The person and company-name passes also use the previous information to identify contexts which indicate the presence of a company name, such as: "ABC stock rose *percentage*". The NE system generates all possible variations for each person and company name it finds; another pass tries to find these variations.

LOUELLA's NE system uses a variety of matching methods. Entities such as dates are found by combining structure format with a list of valid items, i.e. a valid month followed by a number.

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.

In cases where none of the entity parts are in a list of known things, we use surrounding context to identify the name.

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>**

In the above example, LOUELLA does not know what McCann-Erickson is, however, she does know that people are "chairman and chief executive officer of" an organization. Other widely known companies such as "Coca-Cola" are identified through a list of known organizations. This list also helps identify "Coke" as referring to the "Coca-Cola" company.

LOUELLA's NE system makes a basic assumption that any organizations that appear in the headline are the same as, or variations of, the organizations found in the text. Therefore, the system does not examine the headline for organizations until it processes the body of the text. In the previous example, the reference to "McCann-Erickson" in the headline of the walkthrough text is found only after the body of the text is processed during the variation matching phase.

Each variation of a person or organization found is linked to the original name. This link is used in the TE system to identify aliases found in the text for that entity. Only people and named companies found by the NE system will be processed by the TE system.

One additional NE system feature used by the TE system is a "company rename function." If an organization changed—or plans to change—its name, the old or future name is linked to the current name, and the system symbol for the current name is used by all references to either name. The TE system then uses the current name to find all references to the current and old or future names.

The TE System

The TE system builds an object for each organization and person name that contains all of the related information it can find in the article.

An organization object consists of:
1) the organization's name
2) all aliases for that name found in the text
3) one descriptor phrase,
4) the organization type,
5) the organization's locale and
6) country.

A person object consists of:
1) the person's name
2) any aliases for that name in the article, and
3) any titles for that individual which appear in the text.

Much of the information related to the entity name is found during the initial phases of the NE module in the context surrounding the entity name. Appositives, for example, are often good descriptor phrases. The system also links other descriptive phrases and pronouns to the named entity, and these additional descriptions are used to assist the ST system in its information-extraction task. Later, these links will allow for the replacing of noun phrases with, for example, normalized organization template elements.

Once an organization noun-phrase or personal pronoun is identified, the reference resolution module seeks to find its referent. For persons, LOUELLA uses the simple heuristic of assigning the

last person mentioned as the referent, keeping in mind gender constraints. For organizations, the process involves several steps. First, the phrase is checked to make sure it hasn't already been recognized and linked by the NE system. If no match is recognized, a content filter for the phrase is run against a content filtered version of each known organization name; if there is a match, the link is made.

Content Filters:

"the jewelry chain"	=>	(jewelry jewel chain)
"Smith Jewelers"	=>	(smith jewelers jeweler jewel)

For example, if the organization noun phrase "the jewelry chain" is identified, its content filter would be applied to the list of known company names. When it reaches "Smith Jewelers," it will compare the filter against a filtered version of the name. The best match is considered the referent. If there is a tie, file position is considered as a factor, and the closest name is the most likely referent. For generic phrases like "the company," reference is currently determined solely by file position and type.

When a descriptor is linked to an organization name, the syntactic relationship of the descriptor to the organization name is also stored with the phrase. For example, appositives and prenominal phrases recognized by the NE system are tagged with ":APP" and ":PRENOM", respectively. Likewise, references resolved by the reference resolution module are appropriately tagged. The template generator uses a heuristic to choose the descriptor which is most likely correct. The choice is based on a hierarchy which begins with appositives, prenominals, and predicate nominatives, and ends with references resolved by the reference resolution module.

Once an organization or person has been linked to all its variations in the article, the TE system chooses the best name for the element and relegates the rest of the names to the alias category. Assigning the same symbol name to each instance of a template element greatly simplifies the work of the subsequent ST system.

The ST System

The ST system extracts information about complex events that involve template elements like organizations and people. LOUELLA's scenario is about changes in corporate management.

The top–level template of interest is the SUCCESSION_EVENT, which is comprised of:

SUCCESSION_ORG:	an organization template element,
POST:	a string fill,
IN_AND_OUT:	a relational object about each person involved (may be more than one),
VACANCY_REASON:	a set fill.

LOUELLA's strategy is to repeatedly simplify the text before information extraction takes place. This approach allows the most basic elements of the scenario to be identified first. The TE system identifies the primitive template elements (person and organization) involved in a particular scenario. In addition, NE–style methods are applied, at this point, to recognize and tag management–position titles within the text.

The next level of complexity is to find the relational object, IN_AND_OUT. This object is filled by: a pointer to a person template element, the IO_PERSON; a set–fill indicating whether the person is in or out, the NEW_STATUS; a set fill indicating whether the person is currently on the job, the ON_THE_JOB; a pointer to an organization template element representing another corporate entity involved in the change, the OTHER_ORG, (if known); and a set–fill indicating the relationship of the other organization, the REL_OTHER_ORG.

It makes sense to first convert all person template elements into potential IN_AND_OUT objects. In most cases, the sentence clues which will tell the system whether a person is in or out of a position and whether the person is still on the job are also the clues for the succession event itself.

It is preferable, then, to instantiate an empty IN_AND_OUT object around each person element, and then to fill in the rest of the information if an event is extracted.

LOUELLA's ST application consists of three rule packages: ingress.k, which holds all rules for entering corporate posts; egress.k, which holds all rules for leaving corporate posts; and activations.k, which holds all macros for the ingress and egress rule packages.

The Lockheed Martin approach to information extraction is to build sets of floating phrases, i.e. rules, which can glide over each sentence, binding to the right configuration of information. This information is then extracted and reorganized into a lucid account of the events. This approach is similar to the model-based segmentation method used in image-understanding systems. Portions of images are recognized easily, and their configuration is ultimately used to identify the complete image.

By examining a training set of articles for the sentences which report the events of interest, rules are developed. As training progresses, the rules are generalized to cover more and more possible constructs. A typical ingress rule, made up of macros, might look something like this:

```
$subjphr  $conjphr ?IN=$appointvb { $postorg } => c-reassigning-template
```

The binding macros are *$subjphr, $appointvb,* and an optional *$postorg. $conjphr* is a buffer macro allowing the pattern matcher to skip over irrelevant material. This rule contains a variable ?IN which is bound when the rule is matched. This binding is then conveyed to the IN_AND_OUT object as its NEW_STATUS. Other variables are also present within the macro definitions. These variables, when bound, will convey information about VACANCY_REASON, ON_THE_JOB, and the OTHER_ORG to the objects involved in the event. For example, if the ?ACTING, ?IN, and ?FUTURE variables are all bound in a match, then the IN_AND_OUT's NEW_STATUS is IN and ON_THE_JOB is NO, because the text is reporting that the person will be acting in a position.

A difficulty occurs with this method when a sentence identifies a person as leaving one position and entering another. For example, "Judy Jones, president of Exxon, has been hired as CEO of GE." In this case, the person element will have two different NEW_STATUS values, depending on the position being discussed. When this happens, the person element must be re-instantiated as an additional IN_AND_OUT object in order to collect the correct value.

Once an article's information has been extracted, it is then organized into a sensible account based on a model of the domain. This model, along with the final template model (which guides the system's template generator), is constructed at the beginning of training. Both models are based on the scenario specifications.

In the postprocessing stage, we apply any heuristics learned during the course of system development. For this application, the OTHER_ORG portion of the IN_AND_OUT object was filled-in here, based on the information gathered about that person. For example, if at this point LOUELLA knows that a person is leaving one organization and joining another, she can conclude that each organization can be the OTHER_ORG in the IN_AND_OUT object for the other organization's SUCCESSION_EVENT; in effect, the system swaps SUCCESSION_ORGs between succession events to supply their respective IN_AND_OUT objects with OTHER_ORG fillers.

WALKTHROUGH PERFORMANCE

Scenario Template

WARNING: this message does not represent LOUELLA's typical performance. Its F-measure is less than half of LOUELLA's average performance. In fact, this was her next-to-worst score of all the messages. Nonetheless, the system produced the following template for the walkthrough message. The template represents recall of only 15, with precision of 80.

<TEMPLATE-9402240133-1> :=

DOC_NR:	"9402240133"
CONTENT:	<SUCCESSION_EVENT-9402240133-1>

<SUCCESSION_EVENT-9402240133-1> :=

POST:	"chief executive officer"
IN_AND_OUT:	<IN_AND_OUT-9402240133-1>
VACANCY_REASON:	REASSIGNMENT

<IN_AND_OUT-9402240133-1> :=

IO_PERSON:	<PERSON-9402240133-1>
NEW_STATUS:	OUT
ON_THE_JOB:	UNCLEAR

<PERSON-9402240133-1> :=

PER_NAME:	"Robert L. James"
PER_ALIAS:	"James"
PER_TITLE:	"Mr."

Performance on this message reveals two areas in which our system can be improved. First, our method of generalizing had not reached fruition by the time of the evaluation. This message was improved by expanding the definition for one of the floating phrases, i.e. macros, which make up all ingress and egress patterns, and by inserting a buffer phrase into one of the egress patterns.

Adding a buffer phrase allows the pattern matcher to jump over part of the conjunctive phrase in the following sentence:

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.

The following sections show the extraction process taking place. The italicized words represent the rule which is being matched and the variables which are being bound. The buffer addition allows two rules to overlap the sentence, extracting both succession events.

EVALUATION EXTRACTION:

[C-REASSIGNING-TEMPLATE{0} ?IN_AND_OUT=James, 57 years old, is *?OUT=< ?UNCLEAR=?VACANCY_REASON=stepping ?HEAD=down >=?OUT* as *?POST=chief executive officer {0}]* on July 1 and will retire as chairman at the end of the year.

POST-EVALUATION – WITH BUFFER ADDITION:

[C-REASSIGNING-TEMPLATE{1,0} ?IN_AND_OUT=James, 57 years old, is *?OUT=< ?UNCLEAR=?VACANCY_REASON=stepping ?HEAD=down >=?OUT* ***as*** *?POST=chief executive officer {0}]* on July 1 and *?FUTURE=will ?VACANCY_REASON=?OUT=?HEAD=retire* ***as*** *?POST=chairman {0}]* at the end of the year.

The secondary factors that hurt LOUELLA's performance were unsatisfactory post-processing decisions. During development, generic patterns were instantiated to extract organizations which would likely be involved in a succession event. These companies are usually in the act of announcing some event. Then, if succession events are extracted without an organization being directly involved in the event statement, the announcing organization can be tied to the organization-less events. This heuristic worked well; however, no allowance had been made for the case in which a SUCCESSION_EVENT was extracted in the absence of any SUCCESSION_ORG. This was remedied, post-evaluation, by allowing the system to collect all organizations, and to choose an organization during postprocessing to act as a default SUCCESSION_ORG for all organization-less events. Even though post-processing chose the wrong organization for the walk-through message, it still got two extra points for having an organization.

The most striking effect of a deficient post-processing heuristic was the decision to eliminate any succession events which contained only an IN_AND_OUT object, with no other information.

Removal of this heuristic alone, with no other pattern modifications, increased the recall on the walk-through message from 15R/80P to 44R/61P! This is due to the fact that LOUELLA was now producing four succession events, instead of one, each with its own IN_AND_OUT object. This increased the number of correct slots from 8 to 23, even though the additional succession events had no post and no succession organization.

Changes in post-processing, while prompted by performance on the walk-through message, affect system performance as a whole. Consequently, the entire evaluation set was rerun with the changes made to improve the walk-through message. Performance increased from 43R/64P with a 51.63 F-measure to 49R/60P with a 54.04 F-measure.

Of course, the best performance occurs when LOUELLA recognizes all of the organizations present. When improvements were made to the Named Entity task for the walk-through message, the Scenario Template task scores were improved to 65R/56P on that message, for an F-measure of 60.1.

Template Element

LOUELLA produced the following set of template elements for the walk-through message.

```
<ORGANIZATION-9402240133-12> :=
  ORG_NAME:                    "Interpublic Group"
  ORG_TYPE:                    COMPANY
<ORGANIZATION-9402240133-11> :=
  ORG_NAME:                    "McCann"
  ORG_TYPE:                    GOVERNMENT
  ORG_LOCALE:                  McCann CITY
  ORG_COUNTRY:                 United States
<ORGANIZATION-9402240133-10> :=
  ORG_NAME:                    "PaineWebber"
  ORG_TYPE:                    COMPANY
<ORGANIZATION-9402240133-9> :=
  ORG_NAME:                    "Coca-Cola"
  ORG_ALIAS:                   "Coke"
  ORG_TYPE:                    COMPANY
<ORGANIZATION-9402240133-8> :=
  ORG_NAME:                    "Coke"
  ORG_TYPE:                    COMPANY
<ORGANIZATION-9402240133-7> :=
  ORG_NAME:                    "Creative Artists Agency"
  ORG_ALIAS:                   "CAA"
  ORG_DESCRIPTOR:              "the big Hollywood talent agency"
  ORG_TYPE:                    GOVERNMENT
<ORGANIZATION-9402240133-6> :=
  ORG_NAME:                    "WPP Group"
  ORG_TYPE:                    COMPANY
<ORGANIZATION-9402240133-5> :=
  ORG_DESCRIPTOR:              "a hot agency"
  ORG_TYPE:                    COMPANY
<ORGANIZATION-9402240133-4> :=
  ORG_NAME:                    "Ammirati & Puris"
  ORG_DESCRIPTOR:              "a quality operation"
  ORG_TYPE:                    COMPANY
```

<ORGANIZATION-9402240133-3> :=
  ORG_DESCRIPTOR: "one of the largest world-wide agencies"
  ORG_TYPE: COMPANY

<ORGANIZATION-9402240133-2> :=
  ORG_NAME: "New York Yacht Club"
  ORG_TYPE: COMPANY

<ORGANIZATION-9402240133-1> :=
  ORG_NAME: "McCann-Erickson"
  ORG_DESCRIPTOR: "guide Interpublic group"
  ORG_TYPE: COMPANY

<PERSON-9402240133-6> :=
  PER_NAME: "Even Alan Gottesman"
  PER_ALIAS: "Even"

<PERSON-9402240133-5> :=
  PER_NAME: "Peter Kim"

<PERSON-9402240133-4> :=
  PER_NAME: "J. Walter Thompson"

<PERSON-9402240133-3> :=
  PER_NAME: "Martin Puris"

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

<PERSON-9402240133-1> :=
  PER_NAME: "Kevin Goldman"

The score for this document is as follows:

SLOT	POS	ACT	COR	PAR	INC	MIS	SPU	NON	REC	PRE	UND	OVG	ERR	SUB
organization	10	12	10	0	0	0	2	0	100	83	0	17	17	0
name	10	10	6	0	2	2	2	0	60	60	20	20	50	25
alias	3	2	2	0	0	1	0	7	67	100	33	0	33	0
descriptor	3	5	3	0	0	0	2	6	100	60	0	40	40	0
type	10	12	9	0	1	0	2	0	90	75	0	17	25	10
locale	2	1	0	0	0	2	1	8	0	0	100	100	100	0
country	2	1	0	0	0	2	1	8	0	0	100	100	100	0
person	6	7	5	0	1	0	1	0	83	71	0	14	29	17
name	6	7	5	0	1	0	1	0	83	71	0	14	29	17
alias	3	3	2	0	0	1	1	3	67	67	33	33	50	0
title	2	2	2	0	0	0	0	4	100	100	0	0	0	0
TOTAL	41	43	29	0	4	8	10	36	71	67	20	23	43	12

Notice that all of the person objects have actually been extracted. The discrepancy in the score for the person object is due to the incorrect string-fill for the name of Alan Gottesman. LOUELLA incorrectly added the word "Even."

The main improvement to LOUELLA for this walk-through message was the recognition of "McCann" as an alias for "McCann-Erickson," instead of as a location. This allowed the mapping

of the two McCann-Erickson organization objects, which improved our score to 76R/79P from 71R/67P.

Named Entity

Our official NE scores for the walk-through document were 91R/88P. We found two system problems that drastically reduced this score. One problem was the variation for McCann-Erickson, "McCann." LOUELLA threw out the variation because it was known in the gazetteer as a city name. By testing that the variation is part of a hyphenated name, we could then allow the variation to be valid. This one change raised this particular document's score to 96R/93P.

Additionally, LOUELLA found "Even Alan Gottesman" as a person, as well as the variation "Even" later in the document. By forcing LOUELLA to accept the match that starts with a known first name, instead of another part of speech, we threw out this match and raised the document total score to 97R/94P.

With the addition of these two modifications, our total NE F-measure rose to: 94.08.

This document also contains an example of the difficulty in recognizing when a company name is being used as a modifier to a product.

> ... the agency still is dogged by the loss of the key creative assignment for the prestigious **<ENAMEX TYPE="ORGANIZATION">Coca-Cola</ENAMEX>** Classic account.

We are currently looking into expanding the NE module to include a products package. This package will use knowledge about the use of products in text, i.e., how they are referred to and when they include the company name as a premodifier. This type of information may be useful to the analyst who notices a particular person frequently associated with the purchase of certain products, such as Winchester Rifles.

Another interesting ambiguous phrase, which our system did not handle correctly, is:

> 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.

Since it is conceivable that Mr. Dooner could have lost 60 pounds of currency, this makes for an interesting discussion of how smart our systems should be at the named entity level. By possibly making the reference between "reinventing himself" and "lost 60 pounds," the system could throw out the money tag. Another argument could be made that since McCann-Erickson is referred to as "world-wide" in many places, it is even more possible that Mr. Dooner could lose 60 pounds of money. Another possibility is to give our systems the notion of money value vs. weight value; that is, is 60 pounds of currency significant enough to outweigh 60 pounds of weight loss?

AN ANALYSIS OF SYSTEM PERFORMANCE

LOUELLA experienced two bugs during the evaluation which caused at least one document not to be scored in each task. Therefore, we will report two sets of results: the official scores for the incomplete responses, and the unofficial scores for the complete responses which were generated after the bugs were fixed. We consider our true performance to be the complete responses.

Overall, LOUELLA's performance was near the top in all tasks, with F-measures within six percentage points of the top F-measures in Named Entity, within four in Template Element, and within five in Scenario Template.

Named Entity

The Named Entity performance was severely effected by a bug which virtually eliminated one entire response out of the set of thirty; accordingly, the difference in scores between official and

unofficial is most dramatic here. The bottom line scores for Named Entity performance follow, along with the Task Subcategorization Scores for the complete response.

OFFICIAL NE - incomplete (29 texts)

SLOT	POS	ACT	COR	PAR	INC	SPU	MIS	NON	REC	PRE	UND	OVG	ERR	SUB
ALL OBJECTS	2258	2264	2054	0	68	142	136	0	91	91	6	6	14	3

F-MEASURES	P&R	2P&R	P&2R
	90.84	90.77	90.92

UNOFFICIAL NE - complete (30 texts)

SLOT	POS	ACT	COR	PAR	INC	SPU	MIS	NON	REC	PRE	UND	OVG	ERR	SUB
ALL OBJECTS	2276	2296	2128	0	74	94	74	0	93	93	3	4	10	3

F-MEASURES	P&R	2P&R	P&2R
	93.09	92.84	93.33

* * * UNOFFICIAL TASK SUBCATEGORIZATION SCORES * * *

SLOT	POS	ACT	COR	PAR	INC	SPU	MIS	NON	REC	PRE	UND	OVG	ERR	SUB
Enamex:														
organization	449	440	403	0	27	10	19	0	90	92	4	2	12	6
person	373	378	362	0	4	12	7	0	97	96	2	3	6	1
location	110	122	100	0	3	19	7	0	91	82	6	16	22	3
Timex:														
date	113	112	109	0	0	3	4	0	96	97	4	3	6	0
Numex:														
money	76	79	76	0	0	3	0	0	100	96	0	4	4	0
percent	17	17	17	0	0	0	0	0	100	100	0	0	0	0

Note that LOUELLA has achieved near-perfection in four of the six subcategories. It is expected that performance would be even greater over a larger corpus. Since the NE component is a reusable module, it is expected to increase—over time—in recall and precision as it is exercised over a larger corpus.

TemplateElement

Official performance on the Template Element task was degraded by two bugs which caused LOUELLA to lose two articles from the set of 100. Official bottom-line and unofficial total slot scores are:

OFFICIAL TE - incomplete (98 texts)

SLOT	POS	ACT	COR	PAR	INC	MIS	SPU	NON	REC	PRE	UND	OVG	ERR	SUB
ALL OBJECTS	2622	2582	1985	0	207	430	390	1729	76	77	16	15	34	9

F-MEASURES	P&R	2P&R	P&2R
	76.29	76.64	75.94

UNOFFICIAL TE - complete (100 texts)

SLOT	POS	ACT	COR	PAR	INC	MIS	SPU	NON	REC	PRE	UND	OVG	ERR	SUB
organization	589	663	540	0	24	25	99	0	92	81	4	15	22	4
name	539	547	419	0	70	50	58	32	78	77	9	11	30	14
alias	171	158	115	0	1	55	42	364	67	73	32	27	46	1
descriptor	225	239	89	0	52	84	98	289	40	37	37	41	72	37
type	589	663	525	0	39	25	99	0	89	79	4	15	24	7
locale	115	106	67	0	10	38	29	278	58	63	33	27	53	13
country	116	103	75	0	2	39	26	280	65	73	34	25	47	3
person	495	521	465	0	12	18	44	0	94	89	4	8	14	3
name	495	521	453	0	24	18	44	0	92	87	4	8	16	5
alias	170	165	154	0	1	15	10	267	91	93	9	6	14	1
title	166	165	158	0	1	7	6	273	95	96	4	4	8	1
ALL OBJECTS	2586	2667	2055	0	200	331	412	1783	79	77	13	15	31	9

F-MEASURES	P&R	2P&R	P&2R
	78.24	77.52	78.97

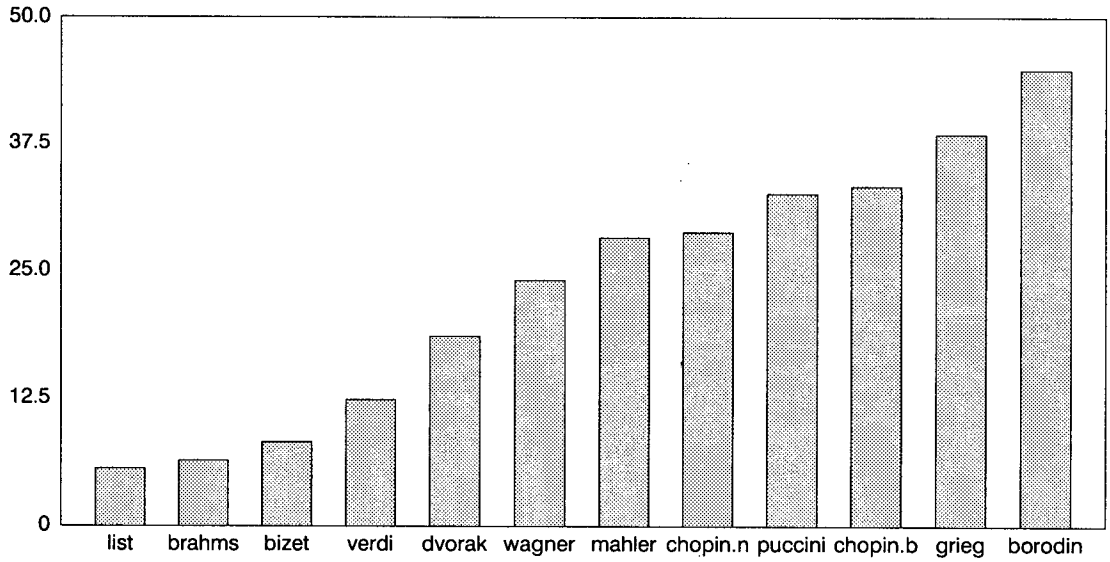

Figure 1: Descriptor F−measures

LOUELLA had very high recall in the Template Element task. She also had very high F-measure for the locale and country slots, and for the descriptor slot. Figures 1 and 2 illustrate F-mea-

sure rankings in the descriptor and locale/country slots, respectively. Since location information is often found in the descriptor phrase, these three slots are somewhat related. High performance on these slots may be due to the attention given to reference resolution during the development of LOUELLA for MUC-6.

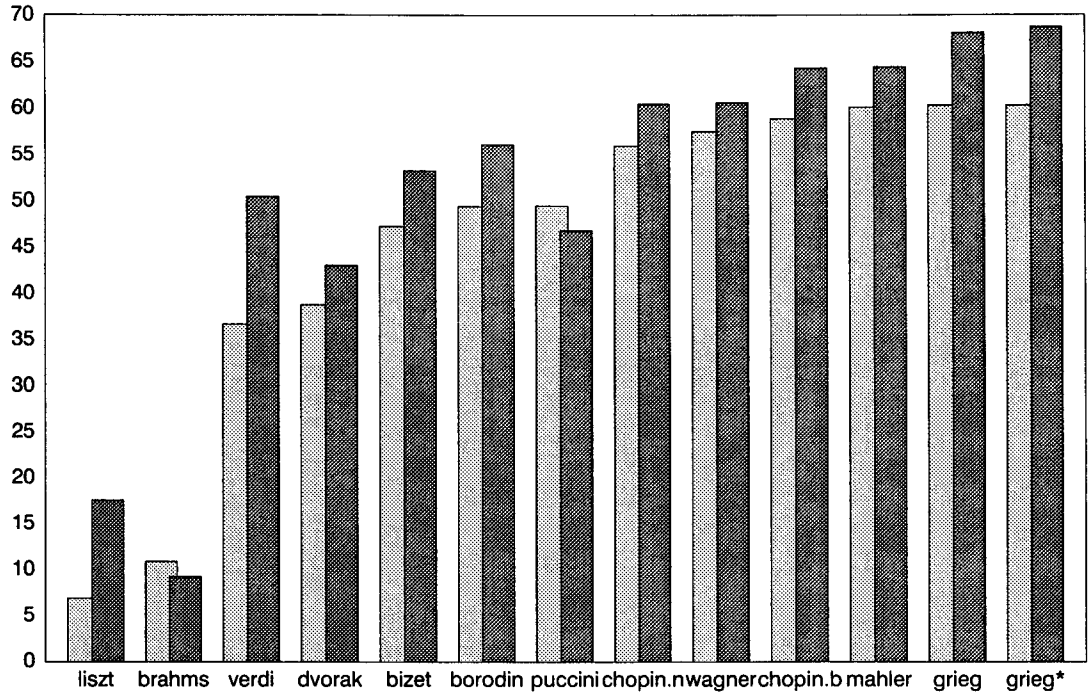

Figure 2: Locale/Country F-measures

A difficulty with the descriptor slot is its mixed role. One function of the slot is to contain any descriptor phrase which is related to an organization's name. This is a true reference resolution task. In addition, however, the slot may also contain a phrase describing an un-named organization. This then requires LOUELLA to differentiate between the two types of phrases and may lead her to overgenerate un-named organization objects, thereby suppressing precision.

Scenario Template

Official performance on the Scenario Template task was degraded by two bugs which caused us to lose two articles from the set of 100. Fortunately, only one of these articles was relevant to the task. Official bottom-line and unofficial total slot scores are:

```
OFFICIAL ST - incomplete (98 texts)
-------------------------------------------------------------------------------
SLOT            POS   ACT| COR  PAR  INC| MIS  SPU  NON|REC PRE UND OVG ERR SUB
-------------------------------------------------------------------------------
ALL OBJECTS    2913  1995| 1267   0  337|1309  391  191| 43  64  45  20  62  21
-------------------------------------------------------------------------------
TEXT FILTER      53    55|   48   0    0|   5    7   40| 91  87   9  13  20   0
-------------------------------------------------------------------------------
                                                    P&R     2P&R     P&2R
F-MEASURES                                        51.63    58.16    46.42
```

UNOFFICIAL ST - complete (100 texts)

SLOT	POS	ACT	COR	PAR	INC	MIS	SPU	NON	REC	PRE	UND	OVG	ERR	SUB
template	53	56	49	0	0	4	7	40	92	88	8	13	18	0
content	196	131	118	0	1	77	12	0	60	90	39	9	43	1
succession_e	196	142	118	0	1	77	23	0	60	83	39	16	46	1
success_or	196	138	87	0	30	79	21	0	44	63	40	15	60	26
post	196	142	75	0	44	77	23	0	38	53	39	16	66	37
in_and_out	262	212	115	0	30	117	67	0	44	54	45	32	65	21
vac_reason	196	142	76	0	43	77	23	0	39	54	39	16	65	36
in_and_out	266	212	170	0	1	95	41	0	64	80	36	19	45	1
io_person	266	212	139	0	32	95	41	0	52	66	36	19	55	19
new_status	266	212	145	0	26	95	41	0	55	68	36	19	53	15
on_the_job	266	212	105	0	66	95	41	0	39	50	36	19	66	39
other_org	189	81	42	0	19	128	20	44	22	52	68	25	80	31
rel_other_org	189	81	30	0	31	128	20	44	16	37	68	25	86	51
organization	115	84	72	0	0	43	12	0	63	86	37	14	43	0
name	112	77	52	0	13	47	12	0	46	68	42	16	58	20
alias	68	56	38	0	1	29	17	16	56	68	43	30	55	3
descriptor	67	31	9	0	10	48	12	15	13	29	72	39	89	53
type	115	84	72	0	0	43	12	0	63	86	37	14	43	0
locale	43	16	13	0	2	28	1	20	30	81	65	6	70	13
country	43	16	14	0	1	28	1	20	33	88	65	6	68	7
person	133	122	98	0	3	32	21	0	74	80	24	17	36	3
name	133	122	93	0	8	32	21	0	70	76	24	17	40	8
alias	85	67	62	0	2	21	3	25	73	93	25	4	30	3
title	81	70	64	0	0	17	6	24	79	91	21	9	26	0
ALL OBJECTS	2969	2102	1349	0	359	1261	394	208	45	64	42	19	60	21
TEXT FILTER	53	56	49	0	0	4	7	40	92	88	8	13	18	0

	P&R	2P&R	P&2R
F-MEASURES	53.20	59.29	48.25

LOUELLA recognized 60% of the succession events after one person-month of development. In fact, she had an F-measure of 69.65 for that slot. This performance shows the system's adaptability. This fact is even more remarkable because of the necessity to write specialized code to handle the peculiarities of this task. Unlike previous extraction tasks in which the event template is built from lower-level relational and primitive elements, this specific task requires that information, such as IN or OUT status, be recognized at the event level but instantiated in the lower level relational element, the IN_AND_OUT object.

TRAINING LOUELLA

Methodology

Ten percent of the 100 development messages were set aside as a blind set for the development phase. This ten percent was chosen based on the size of their keys, so as to accurately represent the complexity of the development set. Over four weeks, the Scenario Template task was able to achieve F-measure of 70.16 on the development set and 52.05 on the blind set. This measure is quite close to our evaluation F-measure of 51.63.

During training, the system is run over both the blind set and the development set of messages overnight, several times a week. Developers can then check the scores at the start of the day and

determine which area of the system is most in need of improvement at that time. This method allows us to check our progress frequently, and to backtrack quickly if a regression is noticed.

Effort

The NE and TE modules of LOUELLA were developed over the Spring and Summer of 1995 by two experienced system developers, one focusing on the NE task and the other on the TE task, with an emphasis on reference resolution. When the evaluation period started, the NE person shifted attention to the TE task, while the TE person shifted to the ST task. Two inexperienced developers were then assigned to the NE task for the evaluation period.

The bulk of the NE effort was directed toward perfecting the rules for recognition. The TE task was more code-intensive because of its reference resolution component, i.e. that task requires an assembling of information gathered up from throughout the article for each organization and person object. The ST effort runs the gamut from domain-specific application design through rule construction and specialized coding; however, the Lockheed Martin NLToolset system provides a basic framework for building an information-extraction application which greatly reduces the amount of effort required. The NE and TE modules themselves are now available for any information-extraction task, and the object-oriented template generator allows the system to easily produce any new template based on the task specifications.

DIRECTION

The reference resolution strategies used for MUC-6 will be expanded to provide more accuracy in identifying related and unrelated organization descriptors, as well as pronoun references. Inclusion of linguistic theory, in addition to other techniques that have been successful for the coreference participants, is a possibility. Research into this area is currently underway.

The procedure for building an extraction system is currently too labor-intensive and haphazard a process, dependent to a great extent on the abilities of the developer. The first step toward remedying this procedure is to build a rigorous syntactic framework which can be used as a template for rule variations. A further step is to investigate the possibility of building a self-training system. Since, at the point of extraction, the system knows a great deal about the components of each sentence, it may be possible to have the system itself generate a set of interesting patterns for a particular domain.

A preliminary effort at linking sub-parts of succession events was attempted for MUC-6. This entailed extracting generic events which were disposable if not linked to task-relevant events. Expansion in this area will include layering of events, as well as an incorporation of time elements, and will ultimately improve the system understanding of the texts being processed.