

MCDONNELL DOUGLAS ELECTRONIC SYSTEMS COMPANY: Description of the INLET System Used for MUC-3

David de Hilster and Amnon Meyers

Advanced Computing Technologies Lab
McDonnell Douglas Electronics Systems Company
1801 East Saint Andrew Place
Santa Ana, California 92705-6520
e-mail: vox@young.mdc.com
phone: (714)566-5956

During the past nine months, our language processing effort has focused on implementing a new NLP system called INLET. INLET relies heavily on the methodology developed for the preceding VOX system. The goal of this work is to produce an NLP shell that can be customized to a variety of tasks and domains.

INLET provides user-friendly graphics-oriented tools for knowledge addition and perusal, to support customization to a variety of domains and tasks. INLET is implemented in C on Sun SPARCstations, in order to support faster analysis than Lisp-based systems and to provide a more widely usable end-product.

At the present time, the basic knowledge addition system is in place, including vocabulary addition interfaces (e.g., Figure 1), a dictionary book tool (Figure 2), a hierarchy editor (Figure 3), and a grammar rule editor (Figure 4). A new Conceptual Grammar has been implemented for INLET, supporting concept hierarchies, a 'conceptual' dictionary, and other knowledge types (such as lexical interrelationships).

A conceptual analyzer, called the *skimmer* (Figure 5), has been implemented to provide robust top-down language analysis capabilities. The skimmer will augment a bottom-up analyzer currently under construction. A skimmer trace is shown in Figure 6.

DESCRIPTION OF THE SKIMMER

The skimmer, implemented in 2 months time, is the only operational component of the INLET analyzer (called LASS, for Language Analysis Support System). Six major processes comprise the skimmer, and are described in turn. The first process performs a pass through the input message to replace all **fixed phrases** with their atomic representation. For example, the literal phrase "farabundo marti national liberation front" is replaced by a concept that represents that phrase. The second process, called the **pre-processor**, locates and parses specialized constructs in the text, such as proper names, number words, temporal and locative phrases, as well as some domain-specific phrases dealing with perpetrators, physical, and human targets. Phrases like "thirty-three", "Ricardo Alfonso Castellar", "three months ago", "San Miguel department", and "member of the FMLN", are all detected and parsed by the pre-processor. Next, a pass through the text detects **key words and phrases**. Words like "attack", "murder", and "bomb", as well as phrases like "set fire to object", "bomb damaged object" are located in the text. Incidents found in this pass are merged or segmented by the **grouping** process. Groups of sentences associated with a single incident are then examined by the **actor/object separation** process to determine perpetrators and targets. The subsequent

slot filling process further examines groups of sentences to fill remaining slots of the MUC3 template. In order to assure self-consistency in a single template, a **semantic trimming and rejection** process trims the templates prior to output. This last process is ad hoc, but substitutes for our current lack of script-based processing.

SAMPLE TEMPLATE

| | |
|----------------------------------|--|
| 0. MESSAGE ID | TST2-MUC3-0048 |
| 1. TEMPLATE ID | 1 |
| 2. DATE OF INCIDENT | - 19 APR 89 |
| 3. TYPE OF INCIDENT | MURDER |
| 4. CATEGORY OF INCIDENT | TERRORIST ACT |
| 5. PERPETRATOR: ID OF INDIV(S) | "GUERRILLAS" |
| 6. PERPETRATOR: ID OF ORG(S) | - |
| 7. PERPETRATOR: CONFIDENCE | - |
| 8. PHYSICAL TARGET: ID(S) | * |
| 9. PHYSICAL TARGET: TOTAL NUM | * |
| 10. PHYSICAL TARGET: TYPE(S) | * |
| 11. HUMAN TARGET: ID(S) | "ROBERTO GARCIA ALVARADO" ("ATTORNEY GENERAL") |
| 12. HUMAN TARGET: TOTAL NUM | 1 |
| 13. HUMAN TARGET: TYPE(S) | CIVILIAN: "ROBERTO GARCIA ALVARADO" |
| 14. TARGET: FOREIGN NATION(S) | - |
| 15. INSTRUMENT: TYPE(S) | - |
| 16. LOCATION OF INCIDENT | EL SALVADOR: SAN SALVADOR (CITY) |
| 17. EFFECT ON PHYSICAL TARGET(S) | * |
| 18. EFFECT ON HUMAN TARGET(S) | * |

A sample of the system's output for message 48 from TST2 is shown above. The system did fairly well, successfully applying an apposition rule in slot 11, for example. However, even though it found the appositive, it incorrectly assigned civilian type to the attorney general, due to an undiagnosed bug. The system often sloughs adjectives, e.g., condensing "urban guerrillas" to "guerrillas" in the perpetrator id slot. Even though the first sentence of message 48 says "...accused the farabundo marti national liberation front (fmln) of the crime", the skimmer failed to fill the organization slot. In this case, failure was due to lack of patterns for accusations. Failure to find the instrument in "...was killed when a bomb placed by urban guerrillas on his vehicle exploded" is due to the absence of a pattern like "bomb placed by actor". Rather, the system knew the pattern "bomb BE placed by actor" where the verb 'to be' is not optional.





