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 preprocessor. 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

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.

		INLET:	Interac	tive Na	atural 1	Languag	je Engi	neering	Tool		
USER: (n	one)	DC	MAIN: (none)	4			• • • • • • • • • • • • • • • • • • • •				
(KD_E∢ € ANALYZI			-	Backup		rune All	nd Find Grow calculate	: guerrilla Grow All Graph	(Options) 6주민 (Dia	Done) ate KB	Quit
	concept	Word :	RY ADDITION				C	Vext Word) (C	ance] (Done)		
8		CONJUGATIONS Present : Grd person : Progressive: Past : Participle : Singular : guerrilla Plural : guerrillas									
EXTEND					SYNTACTIC	FEATURES	5				
			□ adj □ adv	□ aux □ conj	□ det □ modal	afnoun □ prep	□ pro □ quan	□ verb □ wh			
		1 Featur	re : terror	ist/perso	SEMANTIC n/actor/ob						
÷			Figu	ire 1:	Vocabi	ulary	Additio	on Too			







JSER: (none)	DOMAIN: (none)	atural Language Engineering Tool						
[KB Editor]		ionary} [Text loo1] [Options] [SunUS] [Quit						
ANALYZER>	ie.t Tool							
	Text file: nosci	(Save) (Load) (Done)						
	Stat File: reject.tt	Sort: O Off						
	User File: output.tt	Trace: COff						
	Frequency (Context) (Next	Unknown (Add Word) (Skim Step) (Skim) (Print)						
	WORD :	Left: 3 Right: 3 Within: 0 (🕂 🕢						
	Msg: DEV-MUC3-0001 (NOSC)	Left: 3 Right: 3 Within: 0 (A) (U) Processed: Matches:						
	MEANWHILE, THE 3D INFANTRY BRIGADE REPORTED THAT PONCE BATTALION UNITS							
	FOUND THE DECOMPOSED BODY OF A SUBVERSIVE IN LA FINCA HILL, SAN MIGUEL. AN							
	M-16 RIFLE, FIVE GRENADES, AND MATERIAL FOR THE PRODUCTION OF EXPLOSIVES WERE FOUND IN THE SAME PLACE. THE BRIGADE, WHICH IS HEADQUARTERED IN SAN MIGUEL,							
	ADDED THAT THE SEIZURE WAS MADE YES	TERDAY MORNING.						
	NATIONAL GUARD UNITS GUARDING TH	E LAS CANAS BRIDGE, WHICH IS ON THE						
		IS MORNING REPELLED A TERRORIST ATTACK						
		IE ARMED CLASH INVOLVED MORTAR AND RIFLE						
EXTENDER>	C. MESSAGE ID	DEV-MUC3-0001 (NOSC)						
2	1. TEMPLATE ID							
	2. DATE OF INCIDENT 3. TYPE OF INCIDENT	30 DEC 89 Kidnapping						
	4. CATEGORY OF INCIDENT	TERRORIST ACT						
	5. PERPETRATOR: ID OF INDIV(S) 6. PERPETRATOR: ID OF ORG(S)	"TERRORISTS" "FARABUNDO MARTI NATIONAL LIBERATION FRONT"						
	7. PERPETRATOR: CONFIDENCE	REPORTED AS FACT: "FARABUNDO MARTI NATIONAL LIBERATION FR						
	ONT"							
	8. PHYSICAL TARGET: ID(S) 9. PHYSICAL TARGET: TOTAL NUM							
	10. PHYSICAL TARGET: TYPE(S)							
	11. HUMAN TARGET: ID(S)	"PEASANTS"						
	12. HUMAN TARGET: TOTAL NUM 13. HUMAN TARGET: TYPE(S)	PLURAL CTVTLTAN · "PEASANTS"						
	13. HUMAN TARGET: TYPE(S) 14. TARGET: FOREIGN NATION(S)	PLURAL CIVILIAN: "PEASANTS" -						
	13. HUMAN TARGET: TYPE(S) 14. TARGET: FOREIGN NATION(S) 15. INSTRUMENT: TYPE(S)	CIVILIAN: "PEASANTS"						
	13. HUMAN TARGET: TYPE(S) 14. TARGET: FOREIGN NATION(S) 15. INSTRUMENT: TYPE(S) 16. LOCATION OF INCIDENT							
	13. HUMAN TARGET: TYPE(S) 14. TARGET: FOREIGN NATION(S) 15. INSTRUMENT: TYPE(S) 16. LOCATION OF INCIDENT 17. EFFECT ON PHYSICAL TARGET(S) 18. EFFECT ON HUMAN TARGET(S)	CIVILIAN: "PEASANTS"						

	INLET: Interactive Natural Language Engineering Tool
USER: (none)	DOMAIN: (none)
(KB Editor)	Add Vocabulary Dictionary Text fool Options SunOS Quit
Ð	
ANALYZER>	Text file: nosci Text file: nosci Stat File: reject.tt User File: output.tt [Frequency] [Context] [Next Unknown] [Add Word] [Skim Step] [Skim] [Print]
	WORD: Left: 3 Right: 3 Within: 0 ᠿ ↔ Msg: DEV-MUC3-0001 (NOSC) Processed: Matches:
	DEV-HUC3-0001 (NDSC) SAN SALVADOR, 3 JAN 90 [REPORT] [ARMED FORCES PRESS COMMITTEE, COPRETA] [TEXT] THE ARCE BATTALION COMMAND HAS REPORTED THAT ABOUT 50 PEASANTS OF VARIOUS AGES HAVE BEEN KIONAPPED BY TERRORISTS OF THE FARABUNDO MARTI NATIONAL LIBERATION FRONT [FMLN] IN SAN MIGUEL DEPARTMENT. ACCORDING TO THAT GARRISON, THE MASS KIONAPPING TOOK PLACE ON 30 DECEMBER IN SAN LUIS DE LA REINA. THE SOURCE ADDED THAT THE TERRORISTS FORCED THE INOIVIDUALS, WHO YERE TAKEN TO AN UNKKNOW LOCATION, OUT OF THEIR RESIDENCES, PRESUMABLY TO INCORPORATE THEM AGAINST THEIR VILL INTO
EXTENDER>	<pre>Phrase length = 38. Norwilds = 4. Pattern = passive. 1: object 1-13about_number_peasents_of_verfous_ages_ 2: model 3: have 14 have 4: be 15-16been 5: adv 6: ven 17-18 _kidnapped 7: clause 8: by 19-20 _by 9: actor 21-38perpetrator_of_the_perpetrator_[perpetrator]_in_location</pre>
9	Phrase length = 3. Novailds = 2. Pattern = number P: 1-3 _about_ 1: number 4 _number 2: adj 5: country-adj Figure 6: Skimmer Trace