

# Embedded Foreign Language Exploitation Tools

David Day

781-271-2854 • [day@mitre.org](mailto:day@mitre.org)

Army-Contract MOIE



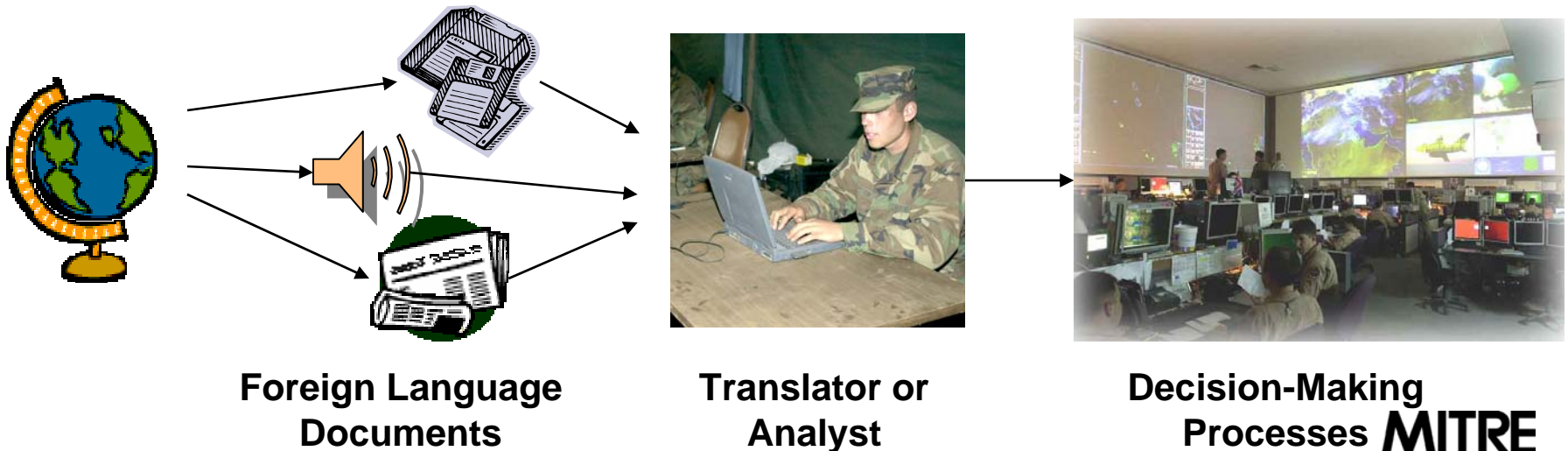
**MITRE**  
Technology  
Program

# Problem

- **Exploiting foreign language material is a key intelligence function, but our ability to access this content in a timely fashion is limited by a shortage of skilled translators and language analysts.**

# Background

- Huge amounts of new and backlogged foreign language material – captured documents and disks, reporting, gray sources, transcribed speech, etc.
- Language analyst skills vary widely, and too often fall short, in foreign and English languages or appreciation of exploitation needs and context



Foreign Language Documents

Translator or Analyst

Decision-Making Processes **MITRE**

# Objective

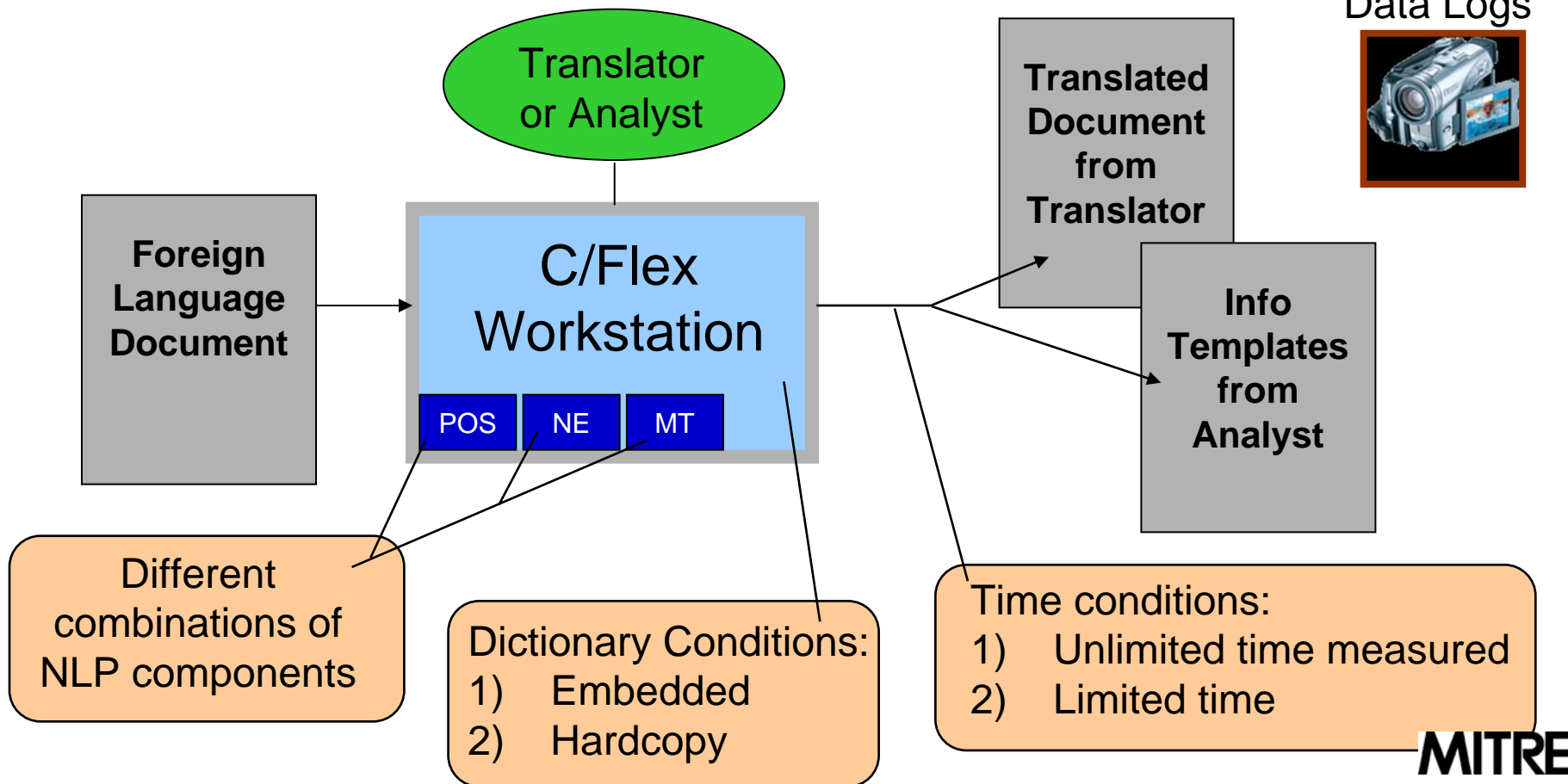
- **Enhance productivity of human exploitation**
  - **Use foreign language natural language processing to boost analyst skill level**
  - **Multiple machine translations as adjunct to translator, to help “triangulate” meaning**
  - **Capture specific names, terms, and metadata that can quickly convey import for “downstream” processing**
- **Instrument tools to improve understanding of usage characteristics, identify successful productivity improvements**

# Activities

- **Construct prototype exploitation tools environment: “C/Flex”**
  - Incorporate machine translation, named entity recognition, standards-based dictionaries, user-editable group-accessible linguistic resources, translation memory, concordances, etc.
  - Integrate with Deployable Harmony DocEx Suite
  - Build toward standards, whenever possible
    - E.g., OLIF2 for multi-lingual dictionaries
- **Perform experiments with language analysts to determine impact of tool suite on exploitation productivity and quality**
  - Use existing metrics (DLIs, DLPT, MT) and others

# Highlight: Experiment Setup

- Controlled studies on usefulness of Flex toolkit
- Compared against subject's normal operating environment
- Multiple experimental conditions



# Demonstration: C/Flex

- Current language focus:

- Chinese

- Next language targets:

- Arabic
- Spanish

- C/Flex implemented in Java 1.4

- Dictionary access via TCP/IP
- Returned XML is displayed via style sheets

- Current MT includes

- Language Weaver
- LEC

- NLP components

- Sentence, part-of-speech and named-entity extraction: Alembic

Callisto - chtb\_226.raw-utf8-minus-sent.sgml.ENAMEX.sgml\*

File Edit Format Tools Help

Main

CB016031.BFJ 404

1994-08-31 12:44:15

迟浩田同坦桑国防部官员会谈

新华社达累斯萨拉姆8月30日电(记者裴善勤)中国国务委员兼国防部长迟浩田和坦桑尼亚国防和国民服务部国务部长基纳纳今天在这里就加强两国和两军之间的合作举行了会谈,并就上述问题达成了一致。

双方对两军30多年来的友好合作表示满意。迟浩田说,中国和坦桑建交30多年来,关系一直很好,两国在政治、经济、军事和文化各个领域里都进行了卓有成效的合作。

MT: LEC

The Xinhua News Agency reaches tired Si-Sa-La-Mu 8 month 3 0 day news( reporter Pei good hardworking) China State Councilor simultaneously Minister of Defense late great field with Tanzania defence with national citizen service department country affair minister Ji fit today in here for reinforcement two countries with between two army cooperation hold talk, and for above-mentioned problem reach consistent.

Both sides for two armies 3 0 long friendly cooperative expression satisfies. Late great field says , Chinese and level mulberry establishing

MT: LanguageWeaver

Xinhua News Agency, tired of Sri Lanka Salam August 30 µs (reporter Pei Shangin) Chinese State Councilor and Defense Minister Chi Haotian and Tanzanian defense and national service department state affair minister » the natrium today on strengthening the two countries and their armed forces' cooperation, held talks, and reached consensus on these issues.

both sides to the two armed forces 30 for many years friendly cooperation with satisfaction. Chi Haotian said that China and Tanzania forged diplomatic ties 30 years ago, bilateral relations

Translation

Xinhua News Agency, tired of Sri Lanka Salam August 30 µs (reporter Pei Shangin) Chinese State Councilor and Defense Minister Chi Haotian and Tanzanian defense and national service department state affair minister » the natrium today on strengthening the two countries and their armed forces' cooperation, held talks, and reached consensus on these issues.

both sides to the two armed forces 30 for many years friendly cooperation with satisfaction. Chi Haotian said that China and Tanzania forged diplomatic ties 30 years ago, bilateral relations

Sentences Gists Machine Translations Dictionary Lookups

Lexemes Paragraphs

Text	POS	Translation
新华社	NT	
达累斯萨拉姆	NS	
8月	T	
30日	T	
电	N	
(	W	
记者	N	
裴善勤	NR	
)	W	
中国	NS	
国务	N	
委员	N	
兼	V	
国防部	NT	
与	A	

Dictionary Lookup

Provider: Group Bi-Lingual Dictionary #2

Entry: 中國

Orthography

Traditional: 中國

Simplified: 中国

Pinyin: ZHONGGUO

Pinyin Tone: ZHONG1 GUO2

Grammar

Syllables: 2

Part Of Speech: NOUN

Definition/Translation

<PL> CHINA; SINO; <General S+T> CHINA, G

Font: 22pt. MingLIU Charset: UTF-8 Task: FLEX

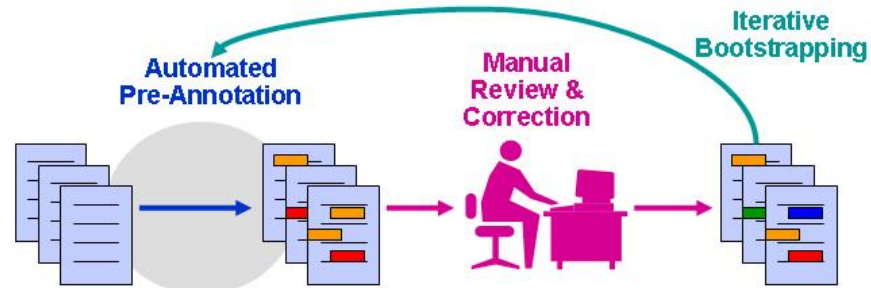
# Impacts

- **Demonstrate utility of automated methods to increase DocEx productivity**
  - **Create initial capabilities that can be tested empirically, including by sponsor organizations**
  - **Collaborate with research and industrial partners**
  - **Evaluate role of natural language processing**
  - **Increase effective DocEx labor pool by boosting non-expert accuracy and speed**
- **Establish use cases for value of interchange standards for language resources**

# Future Plans

- **Expand user-editable resources to include**
  - Entity tracking (specific people, places, etc.)
  - Standardized transliterations
  - Resource tracking to support licensing fees

- **Support mixed-initiative metadata tagging of documents for improved triage and routing**



- **Incorporate and enhance translation memory mechanisms using automatic word clustering**
- **Personalization of displays via XML style sheets, persistent widget layout, ...**