

MUC-7, Tipster 参加報告

福本 淳一† 関根 聡‡ 江里口 善生††

沖電気工業(株) 研究開発本部 関西総合研究所†

fukumoto@kansai.oki.co.jp

Computer Science Department, New York University‡

sekine@cs.nyu.edu

(株) NTT データ 情報科学研究所††

eriguchi@lit.rd.nttdata.co.jp

あらまし

本稿では、テキストからの情報抽出に関する国際会議として米国 DARPA がスポンサーとなっている Message Understanding Conference (MUC) の第 7 回大会の各抽出タスクの概要およびそれらの評価結果について報告する。また、同時に開催された Tipster Phase III 18-month Meeting についても報告する。Tipster においては、情報検索の TREC、要約処理の SUMMAC についての現段階での評価についての報告が行われた。

キーワード

情報抽出, 情報検索, 要約, DARPA, MUC, Tipster, TREC, SUMMAC

Reports on the MUC-7 and Tipster 18-month Meeting

Jun'ichi Fukumoto†, Satoshi Sekine‡ and Yoshio Eriguchi††

Kansai Laboratory, R&D Group, Oki Electric Ind. Co., Ltd.†

Computer Science Department, New York University‡

Laboratory for Information Technology, NTT DADA Corp.††

Abstract

In this paper, we report the results of the 7th Message Understanding Conference (MUC-7) which is supported by DARPA in the States. We also report 18 month meeting of Tipster Text Program Phase III which was held after MUC-7. At the Tipster meeting, there are reports on information retrieval task (TREC) and summarization task (SUMMAC).

Keywords

Information Extraction, Information Retrieval, Summarization, DARPA, MUC, Tipster, TREC, SUMMAC

1 背景

Message Understanding Conference (MUC)[1] は米国防総省をスポンサーとする Tipster プログラムの一環として行われている情報抽出国際会議であり、今回は第7回目となる。今回は、1998.4.29-5.1、ワシントンDC郊外の Hyatt Fair Lakes Hotel で開催された。また、その直後の 1998.5.4-6 に Tipster Phase III の 18 month の meeting[2] も開催され、Tipster の一環で行われている情報検索タスクの TREC、要約処理の SUMMAC の現在の進捗について報告があった。

2 MUC-7

2.1 抽出タスクの概要

MUC への各参加団体は、評価の約半年前に設定される抽出用のタスク定義に従ってそれぞれシステムを開発し、そのシステムを用いて各参加タスクについての評価を行い、その報告を行うものである。設定されたタスクとしては以下のものがあり、TR は今回の MUC で初めて設定されたタスクである。

Named Entity Task (NE)

NE タスクでは、テキスト中の固有名及びそれらの省略表現などの要素を抽出する。抽出する要素は、Entity として人名、組織名、場所名を、Times として日付、時間を、Quantity として金額、割合表現を抽出する。

Template Element Task (TE)

TE タスクでは、Entity(組織名、人名、人工物)と Location についての属性的情報を抽出する。Entity としては、組織(政府関連、一般企業関連、その他の細分類)、人(軍関連、一般市民の細分類)、乗り物(空、地上、海の細分類)といった要素があり、これらの要素の属性的情報として、それを記述または参照する名詞句を抽出する。Location としては、City、Province、Country、Region、Water、Airport、その他といったタイプの情報を認識する。また、各要素に対して Descriptor(要素に関するその性質等の表現を表したも)の情報も抽出する。抽出の際、意味的に同一の要素はひとつの要素として抽出し、それにともない Descriptor 情報もまとめて抽出する。抽出結果は、BNF 記法で表現される。

Template Relation Task (TR)

TR タスクでは、TE で認識された要素 Location(L)、Entity(E) 間の関係を抽出する。関係としては、location_of(L,E)、employee_of(E,E)、product_of(E,E) がある。抽出結果は、TE と同様に BNF 記法で表現される。

Scenario Template Task (ST)

ST タスクでは、シナリオとして設定されたイベントの抽出を行う。MUC-7 で設定されたシナリオは、dry run、training 用として航空機の衝突や事故及びそれに関連するもの(crash events)が設定され、formal run としては打ち上げに関するイベント(launching events)が設定された。抽出する情報としては、事故イベントについては関連した乗り物、場所、時間、事故の犠牲者、事故調査の状況、事故の原因などがある。それに関連する情報として、航空機の出発、到着時刻などを含むフライト情報などもある。

Co-reference Task (CO)

CO タスクでは、テキスト中の代名詞、the+名詞、固有名詞などの名詞のみについての参照関係を抽出する。動詞及び動詞を含むものについては対象外である。参照関係をもつ要素としては、テキスト中で意味的に同一のもので互いに参照しているものがある。抽出された参照関係の情報は NE と同様、SGML タグとしてテキスト中に埋め込まれる。

Multilingual Entity Task (MET-2)

Named Entity タスクの多言語版であり、最終的には日本語と中国語のタスクが設定された。前回の MET-1 から第2回目となるタスクである。タスク定義の最初の段階では、タイ語のタスクも設定予定であったが、最終的には準備不足等の理由で最終的な参加者はなかった。

2.2 評価スケジュール

まず、参加申し込みの後、各参加者に対して抽出結果評価用のタスク定義、正解例となるトレーニング用データがリリースされる。ここでのタスク定義は、最終的な評価用シナリオのコンテキストには依存しない一般的なものであり、シナリオについてもいくつかのサンプルという形式でリリースされる。その後、それぞれのタスクについて、よりしぼられた

形式でのシナリオ、それに基づくタスク定義、トレーニングデータ、スコアラーがリリースされ、各参加者で評価用のシナリオに合わせるようなチューニングが行われる。最後に、一週間の評価期間 (Dry Run, Formal Run がある) の初日に実際の評価対象のテキストがリリースされ、最終日までに抽出結果を送り返し、評価結果および正解がリリースされる。評価結果としては、precision, recall, F-measure の各評価値がある。

今回は、MUC-7 の NE task と MET-2 の Dry Run, Formal Run についてはそれぞれ 97/9 月末、98/3 月始めに設定され、MUC の NE 以外のタスクについては、Dry Run, Formal Run とともに NE の約一ヵ月後に設定された。

2.3 参加団体

最初のエントリーの段階では MUC では 25 団体が参加の意思を示していたが、最終的には 18 団体の参加であった。また、MET では最終的には 5 団体の参加であった。MUC, MET での参加団体のリストを表 2、表 1 に示す。

表 1: MUC-7 参加団体

団体名	NE	TE	TR	ST	CO
American Univ. in Cairo	-	o	o	o	-
BBN	o	o	o	-	-
FACILE(UMIST)	o	o	-	-	-
Isoquest	o	-	-	-	-
Lockheed-Martin	-	o	-	-	-
MITRE	o	-	-	-	-
National Taiwan Univ.	o	-	-	-	-
National Univ. of Singapore	o	-	-	-	-
New York Univ.	o	o	-	o	-
Oki Electric Ind.	o	o	o	o	o
SRA	-	o	o	o	-
TASC	-	-	-	o	-
Univ. of Durham	o	o	-	-	o
Univ. of Edinburgh and Thompson	o	-	-	-	-
Univ. of Manitoba	o	-	-	-	o
Univ. of Pennsylvania	-	-	-	-	o
Univ. of Sheffield	o	o	o	o	o

(表中の太字は日本からの参加団体を示す。)

表 2: MET-2 参加団体

団体名	中国語	日本語
National Taiwan Univ.	o	-
National Univ. of Singapore	o	-
New York Univ.	-	o
NTT-DATA	-	o
Oki Electric Ind.	-	o

(表中の太字は日本からの参加団体を示す。)

2.4 評価結果

MUC-7 では、まず最初に主催者側からの今回設定されたタスクの定義とスコアリングの方法についての説明があり、その後、各タスクでの評価結果の全体的な傾向についての説明が前回の MUC-6 とも比較しながら説明があった。MUC-7 の全体的な報告の後、各参加者からの報告があった。また、スポンサー側として DoD からの発表として情報抽出技術の今後の方向とについてのスポンサー側の考え方について説明があった。そこでは、MUC は Tipster プロジェクトがこの秋で終了するのに伴い、今回の MUC-7 が最後になるということである。また、これまでの評価で非常によい成果が出ており、今後もこの技術はスポンサーとしても重要なものと考えているとのことであった。最後に今後の DoD 側の方針が示され、それに対する参加者からの意見を聞く場も設けられ、これまでの成果や運営方法などについても意見があった。

以下に全体報告で説明のあった各タスクのスコアを示す。

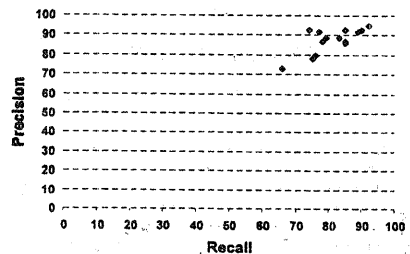


図 1: NE 評価結果

評価の全体的な傾向としては、NE タスクについてはほぼすべての団体はかなり精度で抽出が成功しており、これは前回の MUC-6 と比較してもほぼ同様の傾向が見られるとのことである。したがって、NE

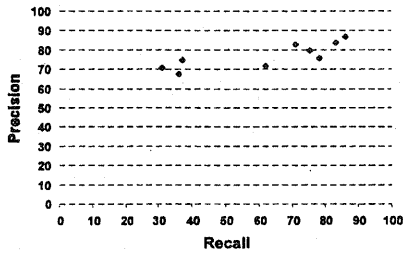


図 2: TE 評価結果

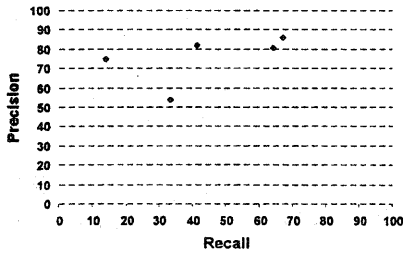


図 3: TR 評価結果

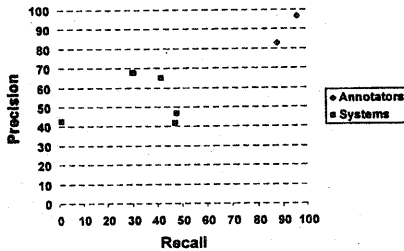


図 4: ST 評価結果

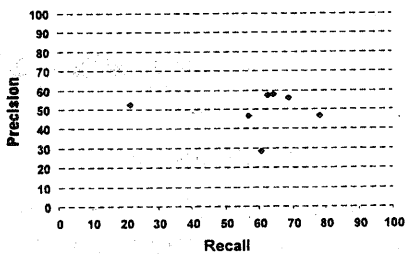


図 5: CO 評価結果

タスクについては基本的な技術としてはほぼ成熟したレベルにあり、今後の方向としては分野依存性の問題をどう解決していくのかに研究の対象が向いているようである。実際いくつかの団体によって学習機構を取り入れた実験が多く行われており、よい成果が出ているとの報告があった。

MUCについては今回で7回目であり、NEレベルについてはある程度ドメインを絞れば(例えば、新聞記事を題材にした場合など)ある程度の性能が出るということが自明となっており、全体的な技術的興味としてはそれをこえるための学習方式に興味の対象が移っていた。学習についてはNEのタスクにとどまらず、CO、TEなどにも積極的に適用が試みられており興味深いものであり、今後の方向づけとなると思われる。スコアについてはすべて公開となり、多少数字的なものへのこだわり(要するに数字レベルでの競争)があるのではないかと思われたが、評価では数字よりも実際にどのような技術を試み、それが結果としてどのようななったのかについて議論が行われ興味深いものであった。

3 Tipster

3.1 概要

本節では、5月4日から6日に行なわれたTipsterについて報告する。Tipsterは、アメリカ政府関係の主催するプロジェクトでテキスト処理に関する技術の推進を目的としている。Tipsterは3期に分けられ、一期(1991-1994)はMUC、TRECの推進、二期(1994-1996)は、それらに加えて、ソフトウェアアーキテクチャの確立、外国語(日本語、中国語)への展開、三期(1996-1998)は、新たにSUMMACの導入があった。3.5節で述べるように、Tipsterは本年秋に終了する予定であり、MUC、TREC、SUMMACについての将来は完全に白紙である。Tipsterにスポンサーされているプロジェクトは主な物に情報検索(TREC)、情報抽出(MUC)、要約(SUMMAC)があるが、他にも多言語処理、ユーザーインターフェース等のプロジェクトのスポンサーもしている。ここで開発された技術は、実際にアプリケーションになっていたり、ユーザーと組んでアプリケーションを指向している(3.4節)。また、Tipsterの終了の宣言と、その後の自然言語に関連したプロジェクトの可能性についての話があった(3.5節)。

MUCの概要については前節で述べたため、以下で

は TREC, SUMMAC の内容について述べる。

3.2 TREC

本節では、TREC について説明する。The Text REtrieval Conference (TREC) は、National Institute of Standards and Technology (NIST) と DARPA をスポンサーとする Tipster プログラムの一環として行われているテキスト検索のコンテストである。TREC の目的は、大規模なテストコレクションと、統一された評価基準と議論の場を提供することで、大規模なテキスト検索アプリケーションの研究を奨励することである。

今年の 11 月に第 7 回目のコンテストである TREC-7 が開催される。Tipster Phase III の 18 month の会議では TIPSTER プログラムのメンバーの中で TREC に参加している TextWise, Inc.、GE R&D、Cornell University、Queens College の 4 団体から報告があった。TREC-7 の Formal Run は、まだ行われていないため、TREC-7 の結果はまだなく、TREC-7 への取り組みを含めた検索技術についての研究経過報告であった。

TREC-7 自体への参加団体は、Tipster のメンバー以外の団体も参加し、全部で 40 程度の団体となっている。日本からも企業を中心にいくつかの団体が参加している。TREC-7 は、必須課題であるメインタスクの Ad Hoc Task と、参加者が参加の可否を選択できるサブタスクがある。サブタスクは、それぞれ検索技術のある部分に焦点をあて、詳しく調べることが目的として設定されている。以下に、タスク名を列挙し、各タスクの概略を示す。

- Ad Hoc Task (メインタスク)
- Filtering Track
- Cross-Language Track
- High Precision Track
- Interactive Track
- Query Track
- Spoken Document Retrieval Track
- Very Large Corpus

Ad Hoc Task は、静的なデータセットから、検索トピックに対する検索結果の精度を評価するタスクで

ある。参加者は Formal Run の半年前に開発用のデータとして 5GB のデータセットと、過去の TREC での 350 の検索トピックと評価セットが配布され、それらを使ってそれぞれのシステムを開発を行う。Formal Run の約 2 カ月前に、TREC-7 用の検索トピックとして新たに 50 の検索トピックが配布される。システムが出力した検索結果をランキングして上位 1000 件を NIST に提出し、提出されたデータに対して recall / precision といった伝統的な評価尺度を使って評価される。

Filtering Track は検索トピックは一定で、新しいデータセットを与えてトピックに適合する文書を識別するタスクである。あらかじめ、その検索トピックに適合する文書がいくつか与えられている。システムは、新しいデータ中の各文書に対し、それが検索トピックに適切なものか否かを判断する。

Cross-Language Track は、言語に関わらず検索トピックに適合する文書を検索できるかを調べることが目的のタスクである。英語とドイツ語とフランス語の 3 か国語の文書があるデータセットに対して、検索を行う。検索トピックは、あらかじめ各国の言語が用意されている。

High Precision User Track は与えられた時間内にどれだけ精度を高めることを目的としたタスクである。5 分間の制限時間内であれば、検索者からのフィードバックをはじめ、あらゆる手段を駆使して検索精度を高めることが許される。

Interactive Track は対話型の検索システムを調べることが目的のタスクである。評価時に検索結果だけでなく、検索プロセス（検索者へのアンケートなど）も対象としていることに特徴がある。

Query Track は検索トピックの表現方法の違いによる影響を調べることが目的のタスクである。参加者は、既存の TREC の検索トピックから、別の形式（自然文、構造化された検索式）といった検索トピックを作成する。これらの検索トピックを使って、検索トピックの表現方法の違いによる影響を調べる。

Spoken Document Retrieval Track は音声原稿の記録を検索するシステムの性能を調べることが目的のタスクである。ニュース放送を人手で書き起こした文書と、音声認識装置で書き起こした文書に対しての検索結果を比較して評価する。

Very Large Corpus Track は非常に大規模なデータベース (100GB) に対する検索システムの性能を調べることが目的のタスクである。検索システムが動作安定性、上位 20 件での適合率、検索時間、検索結

果の作成時間などを評価対象とする。

各団体からの報告で、検索精度をあげる技術のキーは、いかに検索単語を拡張することに集約されていた。いずれも、一度オリジナルの検索語で検索して文書をランキングした後、上位にランクされた文書を利用して、何らかのフィードバックを行い、検索語を増やして再度検索するという手法がとられていた。

Cornell は、最初の検索で上位にランクされた文書を正解データとして Rocchio relevance feedback により検索語を拡張していた。Queens は、最初の検索で上位にランクされた文書から、検索語の関連度が強い単語を相互情報量を用いて決定し、検索語を拡張していた。GE は最初の検索で上位にランクされた文書を要約技術を用いて元の文書の5%に要約し、要約とオリジナルの検索語の関連を調べて、検索語を拡張していた。

3.3 SUMMAC

ここでは、今回新規に設定された要約のコンテスト (SUMMAC) について説明する。

この課題については、初めての試みであるために、評価方法やシステムの開発についても試行錯誤が行なわれた模様である。その過程で得られた経験や知見は、苦勞の後にできたものだけあって色々とバラエティに富んでおり、非常に面白いものであったと思う。

コンテストは以下の課題に分けられ、最初の2つについてはさらに、要約の長さに関して2種類の出力(10%要約、自由長要約)を出す形式で行なわれた。

アドホック	indicative, user driven
カテゴリゼーション	indicative, generic
Q&A	informative, user driven

Indicative というのは、要約した内容が要約前の原文と同じ主題(トピック)を示しているかどうかという点が決め手となっている。この評価方法は、TREC での評価方法に関連している。TREC では、情報検索のトピック(質問)が与えられ、それに対してシステムが出力した記事が、その質問文の内容を表わしているかという事を人間が判断するという方式が取られているが、今回の要約のコンテストでも、要約した文章がその質問で要求された内容に適合しているか、という点で評価が行なわれた。図6にトピックの例を示す。

実際には、元記事が質問文に適合している場合には、要約でも質問文に適合しており、逆に元記事が質

DOMAIN 企業合併

DESCRIPTION 企業合併の発表に関する記事で、その合併に少なくともひとつの日本企業が関連している事。

NARRATIVE 企業合併の発表に関する記事で、その合併に参加する企業の名前が認定できる事。また、合併企業分野、目的などが認定できる事。その合併に少なくともひとつの日本企業が関連している事。

図6: TREC で使われたトピックの例

問文に適合していない場合には要約も質問文に適合していないという事を目的にしている。したがって、実際のコンテストでは、どの場合も一定のトピックに関連する記事と、そうではない記事が混在した形で配布され、評価は再現率 (Recall) と適合率 (Precision) の2つの値とそれらを統合した F-measure が評価の基礎的な値として利用された。また、要約が元記事に対して短い場合には、情報のロスが多くなり、要約が長い場合には、情報のロスが少なくなる事が予想される為、元記事と要約の長さの比を取った圧縮率という考えが導入された。この圧縮率を利用した補正 F-measure というのが、圧縮率と F-measure の積として定義され、この値が実際の議論の中心として利用された。

Informative というのも同様に、要約の内容が元記事の内容の重要部分を保持しているかという事で行なわれたが、その判定のために、検索質問文ではなく、いくつかのチェック項目が用意された。図7にチェック項目の例を示す。

1. Who is the known or suspected hacker accessing a sensitive computer or computer network?
2. How is the hacking accomplished or putatively achieved?
3. Who is the apparent target of the hacker?
4. What did the hacker accomplish once the violation occurred? What was the purpose in performing the violation?
5. What is the time period over which the break-ins were occurring?

図7: Q&A におけるチェック項目の例

この例では対象記事はコンピューターハッカーの

記事としている。これらのチェック項目に対する答が得られるかどうかで、要約の良し悪しが判定される。実際には、得られる/得られないという二段階と、もう一つ中間段階を加えた三段階の二つの方式で評価が行なわれたが、それらには有意な差は見られなかったようである。

User Driven というのは、その記事の内容として求めている物、すなわち、質問文等がシステムに既知の状態では要約を行なう事を意味し、generic というのは、内容が未知である事を意味している。既知である場合には、その質問文の内容を参考に要約が行なえるが、未知の場合には手掛りがないために、パフォーマンスは一般に悪くなる。

また、上記の二つの条件を重ね合せると、Informative+Generic というコンビネーションが存在するが、これには Good Summary という名前が付けられ、将来の課題として扱われていた。

今回の要約のコンテストに参加した団体は以下の16団体である。

- Lexis-Nexis
- Univ. of Surrey
- IBM
- GE CR&D
- SRA
- Intelligent Algorithms
- Center for Information Research
- National Taiwan Univ.
- USC-ISI
- CRL NMSU
- Cornell/SabIR
- TextWise
- British Telecomm
- CIIR UMass
- CGI/CMU
- UPenn

各団体ともそれぞれ、これまでに開発したシステムを利用して要約を行なっていた。以下に利用された技術を簡単に紹介する。それぞれの個々のシステムについて述べるのではなく全体的な流れをまとめ、それぞれ利用された技術を列挙するに留める。当然ながら、利用された技術は幅広く、どれもがこのまとめの流れと同様である訳ではないが、ほぼ、3つのコンポーネントに分けられる。(実際は、3番目のコンポーネントを持つ団体は少なかった。)

1. 前処理

- Term の認定 (形態素情報、品詞情報、固有名情報)
- 質問文中の単語の拡張 (シソーラス利用、コーパスからの自動獲得)
- 照応解析 (Lexical Discourse Structure, Word Cohesion/Chain)

- パラグラフの認定
- テキストのカテゴリの認定

2. 重要文選択

- 単語 (重要語、内容語、質問文中にある単語、ヘッドラインにある単語等) を利用した重要文の選択 (tf*idf 等が利用されている)
- 照応リンク数を利用した重要文の選択
- ヒューリスティック (最初の文、最初の段落、最後の文)

3. 後処理

- 文の短縮化
- 内容的な重複の削除

結果は、補正 F-measure 等が基本的に無記名で発表された。他にも、種々の統計が発表され、その内の一つを図 8 に示す。

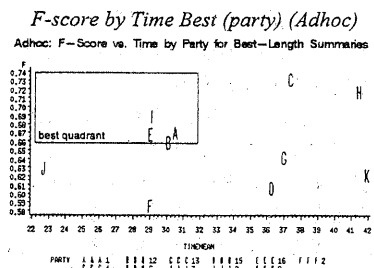


図 8: F-measure と読むのにかかった時間

これは、F-measure と解答を出す読み手が要約を読むのにかかった時間の関係である。Best quadrant の箱の中 (つまり、短時間で読め、かつ内容のロスの少ない要約) が求められるが、図では 4 団体がその範囲にあると認定されている。また、アドホックテストとカテゴリテストを比べた場合、カテゴリテストの方が各団体の結果のちらばりが少なく、ほぼ同様の結果が得られたと報告されている。また、トピックによる要約の困難度の違いも明確に確認されている。Q&A については、コンテスト自身と評価方法もまだ実験的な位置付けにあり、結果に関する考察はあまり見られなかった。

将来への課題や議論としては以下のような話があった。(一部はすでに数団体で試されたり、評価されたりしている。)

- 評価方法(人間の主観的な要約の評価; 良し悪し)
- 人間のパフォーマンスとの関係
- 多言語対応
- 複数文章からの要約
- 他のタイプの文章での要約(ホームページ、電子メール、科学記事、特許)

3.4 ユーザからの声

今回は、かなりの時間がユーザーや潜在的ユーザーからの発表にさかれていた。それらは主に、政府系の情報収集を行なっている人々であった。たとえば、その中の一人は一日に数百の新聞や通信社の麻薬に関する記事を読み、それをまとめ挙げるとい仕事をしている人であった。また、技術の売り込みを行なっている Tipster 関係の人の発表もあり、実際の技術の導入時における困難などの報告があった。現在すでに使われている情報抽出系の技術は主に、MUC のような深い処理に基づいたものではなく、動詞のサブカテゴリの抽出等浅い処理を行ない、それを人間に見せて人間がその後の処理を行なうといった人間との協調システムが多かった。

全体的に、Tipster で開発された技術を使いたいという要望が強く示されていたが、その技術はそのままでは使えず、インターフェースを含めた人間との協調の手法を開発しなければいけないという意見が多かった。また、ユーザーと開発者が一緒にシステムを開発すべきであるという意見も聞かれた。

3.5 Tipster の終了と次期プロジェクト

Tipster は予定を一年繰り上げて、1998 年の秋に終了する。これは、技術的なバックグラウンドというよりも、政治的なバックグラウンドがあるようである。

この会議中に次のプロジェクトとしてスポンサー側(政府側)が考えているプロジェクトへの紹介があった。発表者はしきりに、アイデアを募集しているという発言を繰り返していたので、この内容はまだ流動的なものであると考えられる。このプロジェクトは Tractical Translingual Triage と名付けられている。(発表者は "T cube" と呼んでいた) "Tractical" というのは、完璧でなくてもいいから、実際に使える物を作ろうという事を意味し、"Translingual" からは、文字通り、多言語を強く意識している。"Triage" という

のは、大修館書店の Genius 英和辞書によると 1) (品質による) 選別、2) 治療優先順位の選別 とあるが、NYU のグリッシュマン教授によると、これは、2 の語義に関係した軍隊の用語で以下の意味があるそうである。戦場の病院では負傷者を三段階に分類する。一段階目は、非常に深い痛手を負って、治療しても生存の可能性のない者、二段階目は、治療すれば治る者、三段階目は、治療しなくても、別状はない者。そして、当然ながら、戦場の病院では二段階目の患者だけを治療する。これから類推できるように、次期プロジェクトでは、難しすぎず、簡単すぎないターゲットに絞って展開していこうという話である。具体的な三つの段階は、以下の通りであり、最初の段階を手始めに二番目の段階までやっていこうという意図のようである。

- Situation Analysis: 日常的な情報収集を目的とした処理
- Crisis Analysis: 突発的な情報収集を目的としており、早急のシステムの適応が求められる。
- Battlespace Operation: 戦場で使われる情報収集、伝達を目的とした処理。

二番目のターゲットの具体例としては、ペルーの日本大使館事件等が例として取り上げられていた。シナリオとしては、情報検索、情報抽出を行ない、要約、翻訳の技術も組み込みながら、情報の追跡、収集を行なうというものである。ここでは、Tipster で開発された技術や別の政府系プロジェクトで開発された翻訳の技術などを集約して、役に立つ物を作ろうという意向が伺える。

4 まとめ

本稿では、1998 年春に開催された MUC-7 および Tipster 18-month meeting の概要とそこで行なわれている情報検索タスクの TREC、要約処理の SUMMAC について報告した。

参考文献

- [1] Seventh Message Understanding Conference (MUC-7), DARPA, 1998.
- [2] TIPSTER TEXT Phase III 18-Month Workshop, DARPA, 1998.