

Research Reports of the Research Institute
of Industrial Safety, RIIS-RR-93, 1994
UDC 614.8:681.3

建設工事労働災害事例の発生状況記録中のフリータームの統計分析*

鈴木芳美**

Statistical Quantification Analysis on Free-terms Extracted from Descriptive Statements of the Occupational Accident Report Concerning Construction Works*

by Yoshimi SUZUKI**

Abstract; This study has been made for exploring practical use of descriptive statement of the "Occupational Accident Report". In this study, using information retrieval accident report database system developed in the Research Institute of Industrial Safety (RIIS), 3377 accidents occurred in construction works are selected for statistical quantification analysis on information structure concerning occupational accident.

Procedure of analysis in this study are briefly as follows. Firstly, all free-terms are extracted from the descriptive statements appeared in the item "Accident Situation and Circumstances" by means of the utility program in the database system of RIIS. Secondly, the frequency distribution of these extracted free-terms are examined. Thirdly, high frequency free-terms are selected as key-words for arranging/characterizing information concerning various factors of accidents occurred in construction work sites. In this step, the relationship between these key-words and accident cases is investigated by the statistical quantification analyzing method (quantification method of third type), and using these keywords the possibility for information retrieval on accident cases is also investigated by discriminant analysis. And finally, investigation by cluster analysis on extracted free-terms is made from viewpoint of the relation between these free-terms and other factors such as type of accident and kind of construction work.

Main results of investigations are as follows;

- 1) For all occupational accidents, 39,813 kinds of free-term in total are extracted from the descriptive statements of the item "Accident Situation and Circumstances".
- 2) The frequency of these extracted free-terms are distributed according to the Bradford's law or Zipf's law.
- 3) As a result of examination on actual frequencies of these free-terms, taking "accident type factor" into consideration, the existence of some unevenly distributed free-terms is estimated.
- 4) As a result of cluster analysis of free-terms, good relationship is obtained in dendograms between these free-terms and accident type factor or kind of construction work factor.

*平成 5 年 9 月 8 日第 48 回土木学会年次学術講演会, 1993 年 9 月 29 日第 24 回国際労働安全衛生会議 (FRANCE, Nice) において各々一部について発表

**土木建築研究部 Construction Safety Research Division

Keywords; Occupational accidents, Construction work, Free-term, Information analysis, Statistical analysis, Cluster analysis, Dendrogram.

1. まえがき

労働災害の発生傾向等を分析する際、日常的な災害防止活動等の参考資料として、多くの労働災害事例に関する記録が用いられている。そのため、これらの記録資料を有効かつ適切に活用し得るような体制や条件を整備することも従前より望まれてきている。

しかしながら、労働災害事例を取り扱った原資料（オリジナルデータ）から取得される情報に関しては、これまで情報解析学的なアプローチから分析が行われた例は見受けられない。すなわち、労働災害に関する情報については、情報構造に関する知見や情報内容の性質の把握などが必ずしも充分ではないという側面があった。これらを的確に把握しておくことは、労働災害情報の有効活用に際して、あるいはまた、この種の情報の利用に際して要望されている多様性に対応するためにも、不可欠で基本的な事項と考えられる。

このような観点から、本研究ではこれまでの検討結果¹⁾²⁾を踏まえた上でさらに引き続いて、建設工事労働災害事例に関する記録文書に対して、多変量統計解析手法等を適用した情報解析を行った。特に今回は、前報¹⁾²⁾までに明らかになった労働災害事例に関する情報が有する基本的な構造、すなわち災害タイプや工事種類の差異によって整理される情報構造と労働災害事例の記述にあたって用いられているフリーターム^{*1}との関連を中心に解析を試みた。

2. 分析資料と分析概容

2.1 分析資料

今回の分析に際して用いた資料は、主として前報²⁾と同様、建設工事で発生した労働災害事例に関する災害原因等の調査結果を報告するために作成された「災害調査復命書」と呼ばれる文書資料である。今回はさらに分析資料の一部に「死傷病報告書」を加えた分析を行っている。

前報でも述べたように「災害調査復命書」は、死

^{*1} 人により与えられたり計算機処理により自動抽出される語で、ソースラスや分類体系等によって統制されたり使用制限を受けたりしないものを言う。

亡労働災害あるいは重大災害（一時に3名以上の被災者を伴った災害）などの発生時に所轄の労働基準監督署の係官が現地調査を踏まえて作成するものである。また「死傷病報告書」は、死傷労働災害が発生した場合に災害発生事業所から所轄の労働基準監督署へ提出されるものである。

これらの資料は、労働災害に関する記録資料の中でも、全国的規模でほぼ同一のフォーマットにしたがって作成されるものとして数少ない代表的な資料と言える。またこれらの資料の多くについては、当産業安全研究所において災害事例データベースとしてデータベース化がなされている。

分析対象としたのはこれらの資料に記載される「災害発生状況」に関する記述で、これは100～400字程度で表現には制限のない自由な日本語の文章として記述されたものである。その記述内容は労働災害の発生状況ならびにその要因や原因などに関する事柄が主なものである。

またこれらの記述形態は、建設工事に関する安全や労働災害防止に携わる関係者をはじめ不特定多数の人々が日常的に使用している用語等により構成された文章であるところに特徴がある。したがって、当然のことながら個々の資料については、記述内容の精粗や正確さに“ばらつき”も見られる。

これらの資料の有効活用を念頭に置いた場合には、原資料の記述内容や記述事項あるいは記述形態等の実状に着目し、かつそれらを踏まえた分析が必要と考えられる。そこで今回は、特に建設工事関連の分析事例数を増大させて1985～1988年の4年間に建設工事で発生した労働災害事例（3377件）を分析対象とすることとした。また、一部では建設工事以外の災害事例を含めた場合（1985～1988年、全数では6015件）との比較も行った。

2.2 分析概要

分析はまず災害発生状況の記述中に用いられているフリータームの抽出・切出しを行った後、これらの使用頻度分布状況のチェックを行った。この頻度分布に関しては、各年度ごとあるいは災害タイプごとなどについても、頻度分布状況やその傾向を検討対象とした。

Table 1 Number of free-terms extracted from accident reports.
分析対象事例と切り出されたフリーターム数

年度/ 災害タイプ	全産業 労働災害 (1985~88)	建設工事 (年度別: 全災害タイプ)				建設工事 (災害タイプ別, 1985~1988年)						
		1985	1986	1987	1988	墜落 災害	土砂 崩壊	建設 機械	車両 災害	クレー ン災害	飛來 落下	倒壊 災害
分析事例数 (件)	6,015	732	807	856	982	1,153	183	524	296	271	193	128
フリーターム 総種類数	39,813	8,464	9,059	8,937	9,969	10,478	2,381	6,151	4,209	3,745	3,144	2,231
フリーターム 総のべ数	346,870	36,336	40,285	44,308	48,095	55,619	9,218	25,574	15,298	13,937	10,350	6,296
フリーターム 平均種類数	6.62	11.56	11.22	10.44	10.15	9.09	13.01	11.74	14.22	13.82	16.29	18.44
フリーターム 平均のべ数	57.67	49.64	49.92	51.76	48.98	48.24	50.37	48.81	51.68	51.43	53.63	52.03

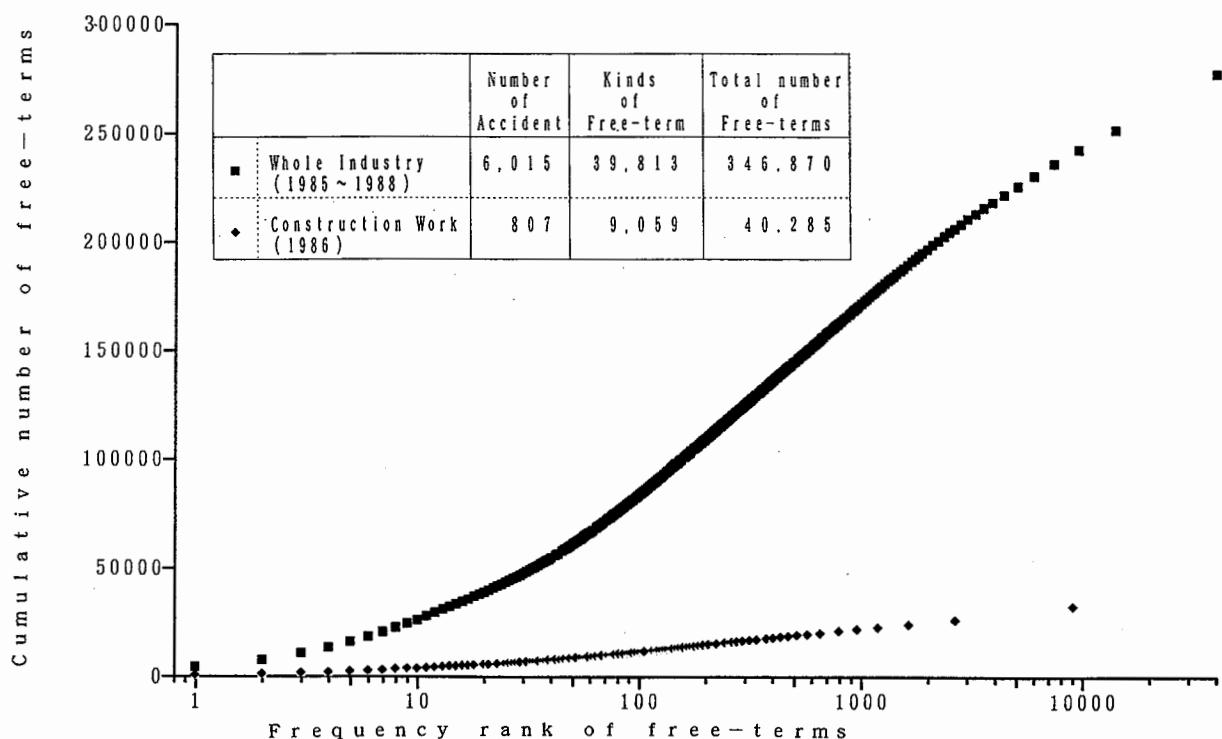


Fig. 1 Braford's distribution of free-terms extracted from accident reports.
切り出されたフリータームに関する頻度順位と累積頻度数の関係 (ブラッドフォードの分布)

さらに、比較的使用頻度の大きなフリータームをパラメータとして選択し、これらを用いて、これまでの分析¹⁾²⁾と同様に、情報構造の把握のための数量化III類を適用した解析、判別分析を適用した事例探索の可能性の検討などを行った。特に今回の分析では、併せて新たに、建設工事で発生した事例について、切り出されたフリーターム相互間の親近性など

に関する分析をクラスター分析等を適用して行った。

3. 分析結果の概略と考察

3.1 切り出されたフリーターム

労働災害の「災害発生状況」に関する記述から切り出されたフリータームの状況を整理すると Table 1

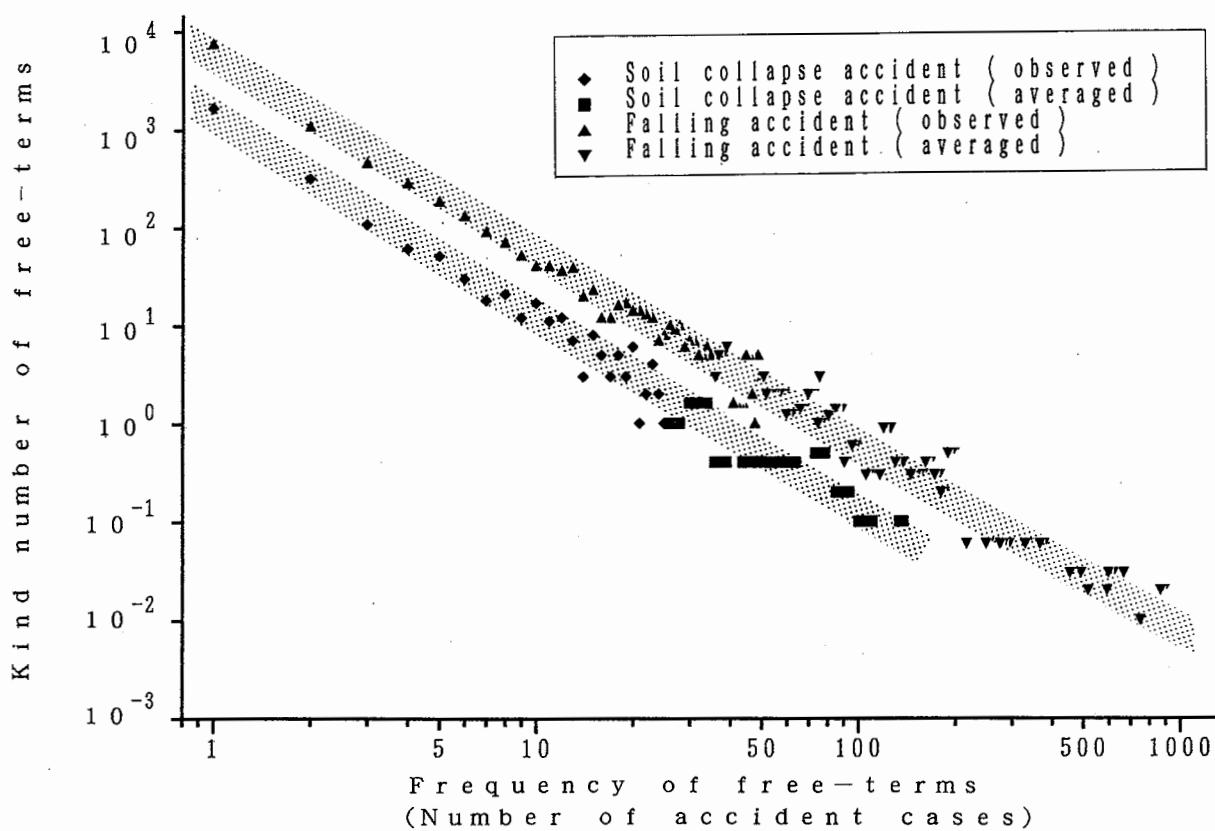


Fig. 2 Frequency distribution of free-terms (Zipf's law).
フリータームの出現頻度と種類数との関係 (ジップの法則)

の通りであった。

これらのフリータームを使用頻度 (当該フリータームの使用事例数) 順に整理 (表略) したところ、これまでに報告¹⁾²⁾³⁾したものと比べ、全体として特別な差異は認められなかった。例えば、抽出された個々のフリータームをその使用頻度順位と累積頻度数との関係で整理してゆくと、Fig. 1 に示すような曲線が得られる。

Fig. 1 の関係は Bradford の法則としてよく知られている分布⁴⁾を示している。すなわち、両者の関係は図の中央部で直線、頻度順位上位 (図左側) で直線上側へ、頻度順位下位 (図右側) で直線下側に各々シフトした曲線を示す。今回のこの分析結果はこれまでに得られた結果¹⁾²⁾³⁾と全く同様で、抽出されたフリータームの頻度分布は典型的な Bradford 曲線を示すことが確認できた。

さらに、年度別や工事種類別あるいは災害タイプ別についても詳細に検討を行ったが、全て同様の頻度分布曲線を示すことが確認された。

その一例として、今回分析された建設工事労働災害事例 (3377 件) を含む全労働災害事例 (6015 件、フリータームの総のべ数 346,870 語、39,813 種類) についての曲線とある 1 年間 (1986 年) に発生した建設工事災害事例 (807 件、フリータームの総のべ数 40,285 語、9,059 種類) に関して得られた曲線とを、Fig. 1 に比較して示した。両者とも典型的な Bradford 曲線を示していることがわかる。

3.2 フリータームの出現頻度分布の傾向

前述したように、抽出された全フリータームの出現頻度 (使用頻度) およびその頻度分布は、Bradford の法則に従うことがわかった。

この関係を別の表現方法、すなわち出現頻度 (当該フリータームを使用している事例数) とある出現頻度を有するフリータームの種類数との関係で見た場合には、両対数グラフ上で傾きがおよそ -2 の直線関係 (Zipf の法則) で示される⁵⁾。

今回の分析結果でも、切り出されたフリーターム

Table 2 List of high frequency free-terms according to accident type.
災害タイプ別に見た出現頻度の大きなフリータームの一覧

災害タイプ	墜落災害	土砂崩壊災害	建設機械災害	車両関連災害	クレーン災害	飛来落下災害	倒壊災害	
分析事例数(件)	1,153	183	524	296	271	193	128	
フリーターム総種類数	10,478	2,381	6,151	4,209	3,745	3,144	2,231	
高出現頻度の フリーターム と その出現頻度数	作業 墜落 工事 死亡 m 被災 者 被災者 下 高さ	893 872 135 757 622 601 94 595 518 494 464	138 135 111 109 101 90 被災者 90 死亡 89 土砂 86 者/現場 79	353 335 被災者 321 318 被災 146 死亡 289 現場 241 運転 228 町 189 m	192 161 154 146 143 119 死亡 122 現場 118 運転 118 町 被災者 112 線 170	187 クレーン 者 170 被災者 146 者 145 工事 126 被災者 118 荷重 115 死亡 84 移動 109 現場 101 式 99	146 被災 者 121 者 118 工事 116 被災者 101 死亡 98 落下 死亡 84 m 80 現場 76 m 69	90 74 73 61 56 50 49 49 49 43

Table 3 List of high frequency free-terms according to year.
年度別に見た出現頻度の大きなフリータームの一覧

年 度	1985	1986	1987	1988				
分析事例数(件)	732	807	856	982				
フリーターム総種類数	8,464	9,059	8,937	9,969				
高出現頻度の フリーターム と その出現頻度数	作業 工事 者 死亡 被災 者 被災者 m 現場 墜落 町	526 458 388 382 372 322 293 284 247 239	作業 工事 被災 者 死亡 m 現場 被災者 墜落 町	559 503 417 401 385 346 343 339 273 243	作業 工事 被災 者 死亡 被災者 m 現場 墜落 町	649 532 478 455 429 381 369 359 319 290	作業 工事 被災 者 死亡 被災者 現場 墜落 町	701 602 572 565 505 490 395 387 371 332

の出現頻度分布に関しては、これまでと同様に、このZipfの法則に従った分布を示すことが確認された。また、年度別あるいは災害タイプ別に見た場合でも同様の傾向を示すことが確認された。

その一例として、Fig. 2には災害タイプ別にこれらの頻度分布状況を示した(図には、土砂崩壊災害および墜落災害の2タイプを代表として表示してある)。図からわかるとおり、頻度分布の傾向には災害タイプの違いによる差異は認められない。なお、本図では、出現頻度の大きなフリータームに関するプロット(図右側)のうち前後に頻度数の欠測値のあるものについては、その欠測値を考慮した平均値をプロットの記号を変えて示してある。

3.3 フリータームの出現頻度の偏奇

全体的なフリータームの頻度分布状況については、前項までに示した通りである。しかしながら、フリータームの中には災害タイプや工事種類の違いにより

その出現頻度に差異を示すものも見受けられる。すなわち、災害タイプや工事種類の違いによって、特徴的な偏奇傾向のあるフリータームが存在する。

このようなフリータームの出現偏奇の傾向をチェックする意味でTable 2には各災害タイプ毎に頻度の大きなフリータームのうち上位10ワードを各々示した。これらの中で、例えばクレーン関連災害における「クレーン」「荷重」、土砂崩壊災害における「掘削」「崩壊」、墜落災害における「墜落」などが災害タイプの違いに基づいて出現頻度に偏奇性の見られるフリータームとみなすことができる。なお、当然のことながら、出現頻度の小さなフリーターム(表中には表示されていない)の中には偏奇傾向の顕著なものが少くない。一方、建設工事における災害について年度別にフリータームの出現頻度の偏奇傾向の有無を示せばTable 3のとおりであり、年度の違いによって出現頻度に大きな差異を示すようなフリータームは見られないことがわかる。

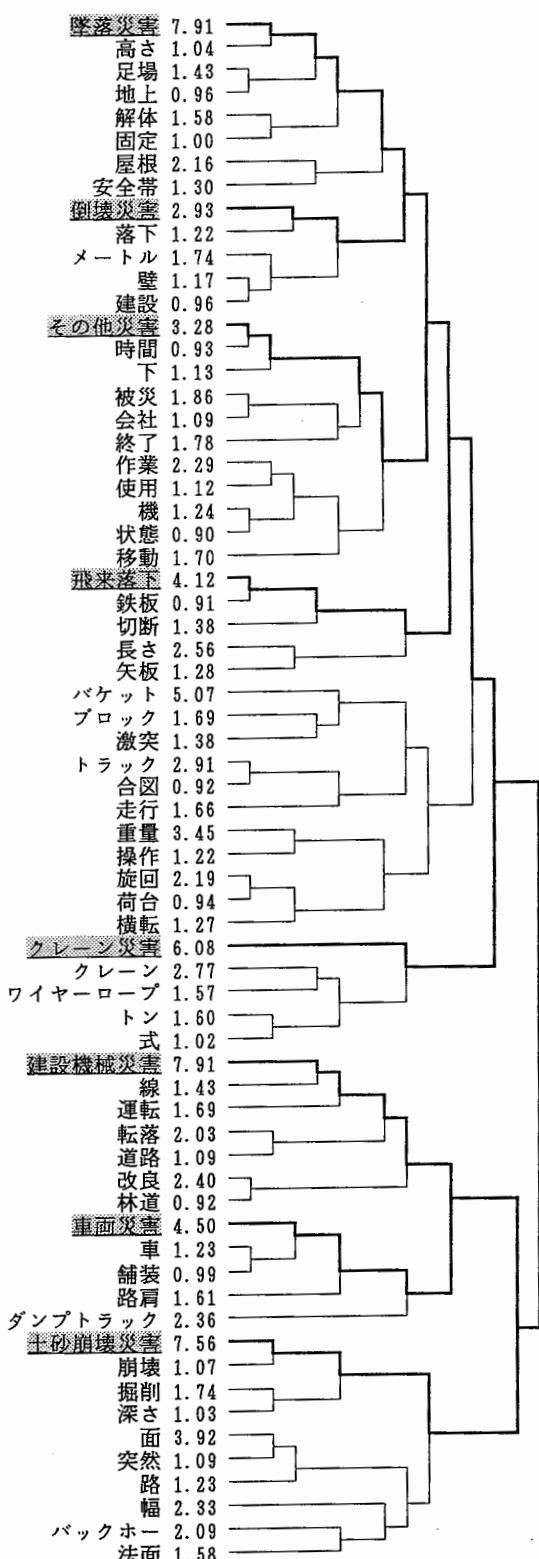


Fig. 3 Relationship between free-terms and accident type (dendrogram: result of cluster analysis).
クラスター分析結果（デンドログラム）による
フリータームと災害タイプとの関連

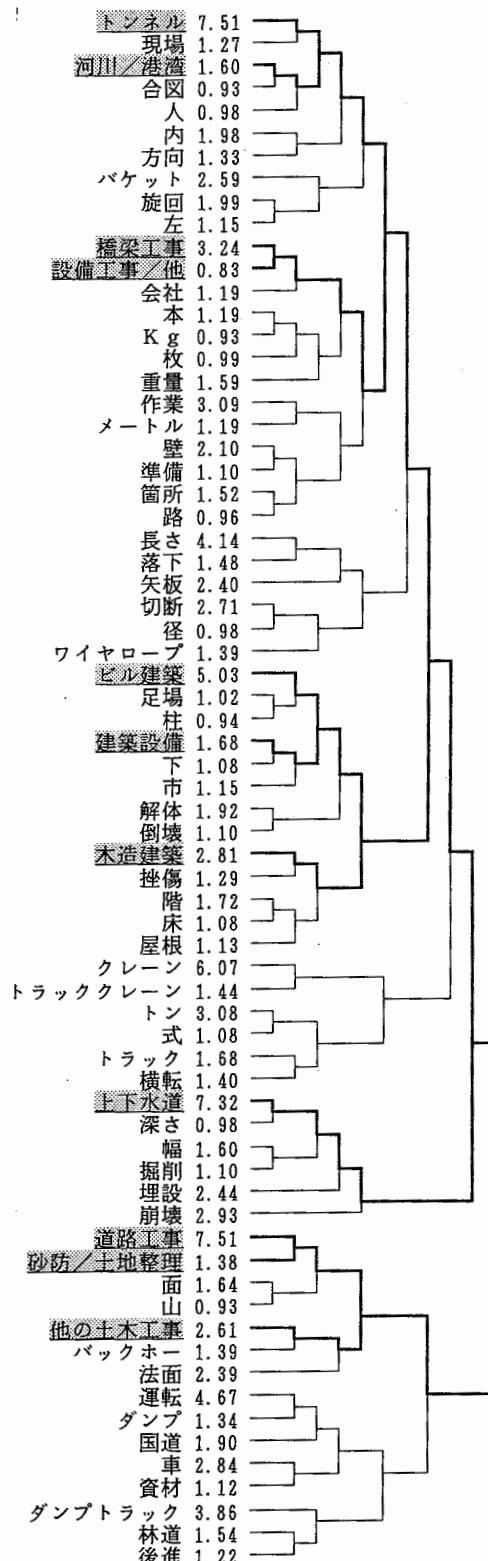


Fig. 4 Relationship between free-terms and kind of construction work (dendrogram: result of cluster analysis).
クラスター分析結果（デンドログラム）による
フリータームと工事種類との関連

これは今回の分析対象が近接した年度の事例であったためとも考えられ、社会背景や時代背景等の大きく異なる災害事例について分析を行って、比較・検討する必要があると思われる。

3.4 フリータームのクラスター分析

3.2 項で述べたように、フリータームの中には、災害タイプや工事種類の違いにより出現頻度に特徴的な差異を示す偏奇傾向のあるものが存在する。このような偏奇傾向をクラスター分析⁶⁾⁷⁾を適用して視覚化することを試みた。各フリータームは各々独立して使用されている訳ではなく、相互に大きな相関を持って使用されていると考えられるため、様々な要因(災害タイプや工事種類など)との関連性を分析し、かつそれを視覚的に捕らえるにはクラスター分析によって得られるデンドログラムが有効と考えられる。

Fig. 3 はその一例であり、災害タイプ(墜落災害、倒壊災害など 8 分類)とフリータームとの関連を、クラスター分析(最遠隣法)により得られるデンドログラムで示したものである。紙面の都合上すべてのフリータームを示してはいないが、この図からわかるように、使用されている各々のフリータームは、一般的に、関連の大きな災害タイプの各クラスターの中に比較的明瞭に区分されて位置している。

同様に、工事種類(トンネル工事、上下水道工事、ビル建築工事など 11 分類)との関係で得たデンドログラムを Fig. 4 に示した。この図では、各フリータームは Fig. 3 に示されたほど明瞭には分類されていないが、例えば建築工事関連、掘削を伴う土木工事関連などのクラスターに属するフリータームを知ることができる。

これらの結果から、一般的にどのような災害タイプあるいは工事種類の災害事例において、どのようなフリータームを用いた記述がなされているか、あるいは、一般的にこれらの災害事例に対して、どのような潜在的意識あるいは共通的認識のもとに労働災害の記録が作成されているかと言った現況を明瞭に知ることができる。また、前報までに述べた事柄、すなわち、労働災害に関する情報構造は災害タイプ等により整理されるということ¹⁾²⁾³⁾が、別の視点からも裏付けられたと考えられる。

4. あとがき

前報²⁾に引き続き、建設工事で発生した労働災害事例

の報告書に記載された災害発生状況の記述に関して情報解析を行った。結果として、フリータームの頻度分布状況が災害タイプ別などについても明らかになつた。また、災害タイプや工事種類に関する情報の抽出や、それに基づいた事例判別の可能性等について、これまでの分析結果¹⁾²⁾³⁾とほぼ同様の結果が確認できた。

本報では、新たに、災害タイプや工事種類の違いと使用されているフリータームとの関連をクラスター分析を通して視覚化することを試みた。その結果得られたデンドログラムにより、各々のフリータームは、災害タイプごとにまとめられるクラスターに比較的明瞭に分類されることが判明した。これらについては、種々の観点からさらに分析を継続する予定である。こうした分析によって得られる知見を総合化することによって、今後も、労働災害事例の記録資料を有効活用するための知識ベースとしての充実を図って行きたいと考えている。

(平成 6 年 5 月 9 日受理)

参考文献

- 1) 鈴木芳美: トンネル工事労働災害情報の性質と構造について、産業安全研究所研究報告 RIIS-RR-91, (1992), 49-58.
- 2) 鈴木芳美: 建設工事労働災害に関するテキスト情報の解析、産業安全研究所研究報告 RIIS-RR-92, (1993), 103-115.
- 3) Y. Suzuki et S. Hanayasu: Statistique analyse de l'information sur les descriptions du compte rendu d'accident du travail concernant travaux publics, Résumés de la 24ème Congrès International de la Santé au Travail (1993), 187.
- 4) 例えは、B.C. Brookes: Theory of the Bradford Law, Journal of Documentation, Vol. 33, No. 3 (1977), 180-209.
- 5) 例えは、小野寺夏生ら: 単純なモデルからの Zipf の法則の導出、第 12 回情報科学技術研究会発表論文集, (1975), 129-138.
- 6) 例えは、田中豊・脇本和昌: 多変量統計解析法、現代数学社 (1983), 230-244.
- 7) 富士通: 統計データ処理パッケージ ANALYST コマンド説明書 (V11 用), (1988), 541-565.