Tetsuo Tomita
Patent Office, Tokyo

Kazutoshi Hattori
The National Diet Library, Tokyo

# Compilation of Thesaurus and Total Index of "Nihon Kagaku-Gijutsu-Shi Taikei" and Analysis of its Documents

Tomita, T., Hattori, K.: **Compilation of thesaurus and total index of "Nihon-Kagaku-Gijutsu-Shi Taikei" and analysis of its documents.**
In: Intern. Classificat. 2 (1975) No. 1, p. 11–21

A thesaurus for the source books "Nihon Kagaku-Gijutsu-Shi Taikei" (History of Science and Technology in Japan from 1850–1960) was compiled comprising 4819 keywords. The characteristics of a thesaurus for historical documents were analysed and names of persons, organizations, laws, geographical places, social tendencies, events and other proper nouns, as well as chronological data were found important and adopted as keywords. Natural and single words were selected from the index of every volume (25 volumes, 32 243 words) and arranged both in a list of loosely grouped facets of classification and in a list according to the order of the Japanese syllabary. After compiling the total index by the thesaurus, 4 627 documents were analysed with the classified keywords. The analysed distributions of keywords well explained the developments of science and engineering in modern Japan. Interesting distributions are those of "military affairs and ordinances", "earthquake disasters", and "names of foreign persons", giving explanations of specialties. Data on the contributions of policies, laws, administrative organizations, research institutes, universities, etc. for science, engineering and daily life including civil and medical activities are explained in figures and keyword-distribution graphs. (Authors)

## 0. Preliminary remarks

In 1960, the History of Science Society of Japan commenced planning of a series of books on history of science and technology in Japan since 1850, in commemoration of its twentieth anniversary. The series was entitled "Nihon Kagaku-Gijutsu-Shi Taikei" (History of Science and Technology in Japan : here after reference as "Taikei"), and about one hundred Japanese historians of science and technology participated in the work.

Publication of the 25 volume series, by Dai-Ichi-Hoki Shuppan Co. Ltd, commenced in March 1964 and ended in August 1970. In March 1968, Professor Mitsutomo Yuasa proposed the preparation of a thesaurus of the history of science and technology in the form of a total index covering all 25 volumes, to be published as a supplementary volume. After consulation with members of the History of Science Society of Japan, he appointed us to prepare the total index in a cross index style. The supplementary volume was published in January 1972. The Nihon Kagaku-Gijutsu Kenshujo (The Computer Centre, Union of Japanese Science and Engineering : hereafter referred to as JUSE) voluntarily collaborated with us in the programming and running of an electronic computer.

We express our thanks to Dai-Ichi-Hoki Shuppan Co. Ltd. and JUSE company, who assisted in programming and running the electronic computer. We also express our thanks to Dr. Diemer, professor of Düsseldorf Universität, who advised us to make this contribution at the XIV International Congress of History of Science in Tokyo.

## 1. Introduction to the "Taikei"

The "Taikei" consists of 25 volumes and covers the history of all fields of science and engineering from about 1850 to 1950. A title list of the Taikei is shown in Table 1. Each volume was edited by the cooperative efforts of several members, for instance, volume 14 "Astronomy and Earth Science" was edited by three groups i. e. scientists and historians of astronomy, those of climatology and oceanology, and those of geography and cartography totalling 9 persons in all. Alltogether, about one hundred persons participated in the 25 volume work. They gathered very many historical documents, analysed and compared them, selected important parts of them, and arranged and explained them according to developments in science and engineering. Each volume contained max. 349, min. 152, average 218 documents and their explanations, and has max. 614, min. 538, average 564 pages. An index is also included at the end of each volume, containing max. 1834, min. 533, average 1290 words. We will say more about the index in 2.2.

## 2. Compilation of thesaurus and total index of the Taikei

### 2.1 Special characteristics of a thesaurus of historical documents

Recently, we have seen many kinds of thesauri for science, medicine, engineering, management and so on. But, to the best of our knowledge, there have been few trials to make classifications or thesauri for historical documents. Let us consider some characteristics of historical documents. Historical documents are similar to the experimental or the observable data used in pure sciences. They differ from scientific papers or new items. Such data must be processed before they yield useful information. The characteristics of historical documents include:

1) The contents of documents are limited to special events or special circumstances in particular areas and, therefore, they have no universal validity.
2) As each document is independent and isolated from other documents, one document cannot be used in place of another. This characteristic rejects the making of abstracts of historical documents.
3) Some historical documents are valuable not because of their contents but because of the circumstances under which they were made or the influences they exerted on society.
4) Documents relating to any social problems may also be historical documents of science and technology.
5) The form, style, language and other bibliographical data of documents may be important to historians, as for example, the first editions of text books in national languages or architectural illustrations.

These characteristics of historical documents have a bearing on thesauri as follows:

A. Proper nouns such as names of persons, institutes, universities, companies, congresses, acts, and treaties are necessary Key-words.
B. Famous (notorious) events, e. g., large earthquakes and great fires, or social tendencies, e. g., "Buy-Japanese-Goods movement"; "Wealth and armament", are important Key-words.
C. Chronological and geographical data are necessary.
D. The form or style of documents, e. g., language, style of printing, are relevant in many documents such as diaries, essays, autobiographies, official documents, statistical data, graphs, maps, pictures, and blueprints.
E. Terms in all fields are necessary because the various fields in which historical documents are related to developments in science and technology are unlimited.

### 2.2 Design of thesaurus

Generally, in preparing a thesaurus, analysis of all documents is seldom achieved since, in many cases, the number of documents to be analysed increases with time, resulting in a large accumulation in future. Thus, one usually prepares a thesaurus with old documents, supplementing it with new terms later. In our case, however, all documents for analysis have either been printed or are being prepared for printing. Therefore there will be no future accumulation of documents. Moreover, there will be no opportunity to maintain the thesaurus after publication. In addition, no arbitrary or random characteristic, which would indicate statistically currents and trends of history, can be ascertained from Taikei since it was compiled from selected or typical documents only. Therefore, the method of sample analysis of documents, was hardly appropriate. This pointed to the need to analyse all documents, a task requiring the investigation of 14,088 pages, therefore, we considered using the index included at the end of every volume as a rough indication of the content of the documents in the volume. The number of terms appearing in each index is given in Table 1.

Such index contains min. 533, max. 1834, average 1290 words, the total for the 25 volumes being 32,243 words. This means that 8.92 words were extracted from a document, or 2.28 words from a page. But since some words in an index may correspond to several pages, we checked the average frequency of a word in Volume 1 and found that a word is extracted from 1.29 places. Therefore the number of words extracted from all 25 volumes is 40.500 or 2.88 word per page, 7.45 words per document. We concluded that the words of these indices had suitable depth as resources of Key-words. Further, these were regarded as valuable guides in compiling the thesaurus owing to the following factors;

1) Words given in the index at the end of every volume are natural words.
2) It is certain that those exist in some parts of the volume.
3) It is supposed the editors of each volume thought those words somewhat important.
4) As every volume has an index at the end, we can see which words are important in the 25 volumes as a whole without twice analysing the documents.
5) Each Index has deviation and bias of picking up words, but we can cancel such deviation by amassing the words of the 25 indices, since the deviation or absence of common words of a certain volume may be corrected by other volumes. If words are not supplied, their absence justly indicates the lack of documents in a particular field. It is unnecessary to supply such words in our work.

Total numbers of words in the indices of all volumes is 32,243. Amongst them are personal names. The number of personal names in an index is min. 116, max. 681, average 681 and total 8,655 (net number of persons is 6,141); amongst these, the number of foreign names is min. 0, max. 461, average 63.4 and total 1,584. At first personal names were selected according to frequency. The names of other persons which appeared more than once as writers of documents, were added, as considered important. We got 897 Japanese names and 146 foreign names. After extracting the personal names, 23,588 words relating to subjects remained. These words were classified and necessary terms were extracted in nine stages.

Attention was paid to the following points:

1) When we sorted words, we took vertically crossing divisions of fields, in contrast to the divisions of the 25 volumes. We also took various vertically crossing facets in classifying words during the assortments.
2) Currents or trends of developments of society, i. e., historical trends, are often changed by unpredicted or accidental events, such as wars, revolutions, earthquakes, typhoons and floods, and epidemic diseases. In China and Korea, historians paid attention to such unusual events as premonitory symptoms of social perplexity. Therefore as will as myths we devised a special facet entitled "Unusual events and phenomena."
3) Many kinds of proper nouns are necessary. Besides personal names, we took the following proper nouns as facets: academic societies, research institutes, research organizations, international organizations, congresses, governmental organizations, companies, societies, labour unions, acts, treaties, objects of policies and social movements, social tendencies, universities, schools libraries, museums, hospitals, important edifices, journals and newspapers.

4) To construct the thesaurus, we combined different facets to construct a field or group, for instance, for the classification of contents of researches, we combined the three facets "branches of learning", "concepts and objects of research", and "experimental techniques".

Further, our thesaurus has the following distinctive points:

5) One part of the classification system was based on names of places such as continents, oceans, seas, countries, prefectures, important cities, mountains, rivers and lakes. Since borders of nations change because of war, we used the borders of 1940 as indicators of Key-words.
6) In chronological data, we took three different sets of Key-words: Key-words indicating era, those indicating each year since 1800 A. D. and those indicating five-year intervals.
7) We prepared a facet for styles or kinds of documents. Key-words of the facet are also included in other facets, for instance, "abstracts of papers" is also classified under "tools of sciences" and "publications".

Finally we obtained three term lists; a classified Key-word, a Key-word list in the order of Japanese syllabary, and a list of personal names in the order of Japanese syllabary.

The classified Key-word list consists of seven main divisions and four auxiliary divisions, they are.

Main Divisions

1) Science, Research (Medical Science)
2) Engineering, Industry
3) Policy, Economy
4) Social Problems, Labour Problems
5) Lives, Medical Treatment, Education (City)
6) Thought, Philosophy, Religion, Fine Arts
7) Unusual Events, Catastrophes or Phenomena

Auxiliary Divisions

8) Names of Places
9) Styles, Forms of Documents (or Artifacts)
10) Ages
11) Names of Persons (Japanese and Foreigners)

In the seventh bough of the classification, "Unusual Events, Catastrophes or Phenomena", we included 84 Key-words of calamities: typhoons, earthquake disasters, poor harvests, fires, mineral poisoning, air pollution, disease, wars, and several kinds of events such as revolutions, panics, riots, genocides, clairvoyances and so on. Included are the four largest typhoons, four great earthquakes and ten wars.

As the Key-words consist not of sentences, but of single or compound nouns, hierarchical relations between words are not strictly logical. An item in a hierarchy is not a word including all terms, but a noteworthy or representative word. Sometimes we took plural terms as headings of groups or items.

Terms having hierarchical relations are regarded as Broader Terms or Narrower Terms, and terms of same rank are regarded as Related Terms.

Some of the Key-words were omitted during the process of making the total index, the thesaurus finally containing 3,772 Key-words of subjects, 1,047 personal names and chronological data.

### 2.3 System design

According to our system, each document is first recorded in the computer memory, then the computor arranges each document in a file-order. This file is output not only on lineprinter in list form, but also on paper tape in punched form together with type picking code for monotype setting. These elements are described in detail below:

1) Key-word coding and parity check

In general, it is very difficult to input a data such as Key-words of this index which are written in Chinese characters. To solve this difficulty, we developed a process by which a Key-word written in Chinese characters is converted to a number. For this, we developed a parity-check to guard against errors in analyzing, punching and handwriting. The parity check is achieved by writing the initial letter of the Key-word expressed in the Hepburnian system of romanization.

2) Compilation of dictionaries

Two dictionaries were used as a manual of computer input. One dictionary consists of key-word terms arranged in the order of the Japanese syllabary; the other consists of key-word terms arranged as elements in a facet or multi-dimensional classification.

3) Refiling

There are two steps to produce a refiling, that is:

$$KW \left\{ \begin{matrix} MOVE \\ \longrightarrow \\ DUPL \end{matrix} \right\} KW \left\{ \begin{matrix} MOVE \\ \longrightarrow \\ DELT \end{matrix} \right\} KW$$

First Step Second Step

4) Out-put tape and format for type setting.

5) Flow chart in operation.

The program consists of five sub-programs named as SAKUIN 1–5, with 2,200 job-steps. Program language is COBOL, and the Computer used is TOSBAC 3400/41.

### 2.4 Analysers and their consistency

A large number of analysers were necessary to prepare an index as comprehensive as ours. They had to be graduates of a college of technology, thus having a basic knowledge of science and technology and they should also be trained librarians, having some experience of indexing and classification. We decided to assign two volumes per analyser.

Since analysers have different backgrounds and experience, the analysed results may differ widely. To guard against such differences and to measure personal error we devised a consistency check for all analysers.

There are two methods of devising a consistency check. The first is a statistical method, which examines the key-words selected when the same document is distributed to all analysers. Key-words selected by many analysers are more valuable, than those selected by few or by no analysers. Those selected by few or no analysers are referred to as "Noise."

Arranging the selected key-words in the order of frequency of selection we set a boundary line at the key-word where the summed frequency of the key-words of the highest frequency reaches medium of the total. By this method, the range of high frequency key-words is too broad. Therefore we need a more rigorous method of check consistency.

The second method, which is a complementary method, investigates how many key-words selected by an analyser match the key-words we have already selected as correct key-words for contents of the documents. Then we count matching key-words selected by each analyser.

We recognized that both statistic and complementary counts are not based on how many key-words are selected by an analyser, but how many correct key-words are matched according to their choice.

There are some tendencies or characteristics of individual analysers which affect the number of key-words selected and the number of key-words matched thus affecting the total number of key-words. This tendency is indicated in Fig. 3.

Among Key-words we found Key-words of personal names had 100 percent consistency. We must pay much attention to proper nouns. After the examination of the consistency check, we asked analysers to analyse each of the volume. The average number of Key-words per document was 7.8.

### 2.5 Duration

The work of compiling the thesaurus began on 9th October, 1968 and we spent 3–4 hours of voluntary work per evening every week, with 2–4 members attending a meeting. It took 232 person's evenings to make the thesaurus, excluding simple tasks such as rearranging words in the order of Japanese syllabary, making fair copies and making duplications. The classified Key-word list was completed on 23rd January 1971 and we had several meetings to explain and amend Key-word lists to analysers from 6th March 1971.

When the analysers' work was completed, the data for indexing documents were handed over to JUSE company. The computer of TOSBAC 3400–41 ran for $10\frac{2}{3}$ hours and a monotype machine of Dai-Ichi-Hoki Shuppan Co., Ltd. set up types for nearly 60 hours.

## 3. Certain analyses of history of science and engineering in Japan

### 3.1 General discussion

Having finished the publication of the supplementary volume, we tried to use the data of the index.

We knew some researchers counted yearly numbers of documents (treatises, papers and so on) on certain subjects to account for some phenomena, and we were able to take a further step forward in this total index.

We were easily able to get chronological distributions of Key-words or sets of Key-words in addition to distributions of documents of Taikei because they were analysed and indexed with Key-words.

We built up a system to exploit related documents with a set of Key-words for a certain theme, to sort the documents in the order of chronological data and to classify all Key-words of the exploited documents according to a classification.

Before discussing the result of our analysis, we must make the limitations of our system clear.

1) The Taikei consists of 25 volumes of which documents are selected by editors of each volume. In a volume, there may be only one typical document for a certain event or a certain field. In addition, there would be many kinds of editors' fondness in selecting documents. We can hardly think the documents of all 25 volumes are randomly sampled from the document-population of history of science in Japan.

As a result, we considered there were 25 elements of sets of documents independent of each other, and selected our themes of analysis as fully macroscopic to cover all 25 volumes.

2) Some documents had plural Key-words of chronology, eg. a year of a discovery, a year of the first report and a year of publication of a treatise, or a year of an event and a year when a memory of the event is published. In such cases, a computer could not distinguish the meaning of all data of years, and we had to choose either to count such a document as plural or count it once. We chose to count a document once at the earliest of chronological data.

Further, we gathered documents into five-year intervals as a set of intervals of 1800–1804, 1805–1809 A. D. and so on.

3) Documents of Volume 7 "International" and those of Volumes 8, 9 and 10 "Education" were omitted because their documents introduce noises in our analysis of history of science. The documents of "International" were of foreign countries and the documents contained in "Education" had similar Key-words to science such as mathematics, physics and so on. We analysed 4627 documents over 21 volumes, omitting these 4 volumes.

We selcted 26 themes and picked up Key-words related to each theme. We also took tue seven main divisions of the thesaurus and added names of Japanese and foreign persons so as to classify the Key-words obtained from exploited documents of a theme. Finally we obtained 260 time sequencial data including numbers of documents. Indicators for *science and research* are comparatively high because the classification bough under the title contains the largest number of Key-words among seven classification boughs.
Several examples are shown in the following sections.

### 3.2 General tendencies

We show the distributions of Key-words from all 4627 documents in Fig. 3A.

Roughly speaking, modern Japanese sciences and engineering rose about 1840 with the learning of medical and military sciences through the Dutch language, after the western impact on China, especially after the Opium War in 1838.

The data, of course, tell us little prior to 1853, when Commodore Perry came to Japan, and after 1955, we find a sharp rise in the field of science, engineering, policy, and so on, in the interval 1850–54.

European sciences and engineering poured into Japan after the Meiji Restoration (1868) as a floodgate of national isolation collapsed. A peak from 1850 to 1900 means a rushed influx of science and engineering into Japan. The flow stopped temporally during the Sino-Japanese War in 1894–95 and Russo-Japanese War in 1904–05. There rose national sciences and engineering, especially after World War I, which prospered until 1945.

During the second peak from 1910 to 1945, serious damage was done to the development of science and engineering, due to the great Kanto-earthquake which struck Tokyo, Yokohama and surrounding industrial zones on the 1st of September 1923. Nearly seven hundred thousand houses collapsed or burned and ninety seven thousand persons were killed.

When Japan opened fire in China, mobilizion of science and engineering brought a noticeable high peak in the distribution of Key-words classified for *science and research, engineering and industry,* and *policy and economy.*

Only the indicators of *policy and economy*, and *social and labour problems* had high points after the defeat in World War II. They reflect the U. S. occupation and social perplexities.

Apart from Fig. 3A, we show distributions of Key-words of documents concerning "Science and Research" in Fig. 3B. All Key-words of the first bough of the classification were listed and all documents concerning them were analysed. We also show the distributions of Key-words of documents concerning "Engineering and Industry" in Fig. 3 C, which corresponds to the second bough of the classification.

Comparing Fig. 3 A, 3 B, and 3 C, general features are similar, but we find some different behaviour in the period between 1890 and 1909 and the era after the end of World War II. *Science and research* in Fig. 3 B fall during 1890 and 1909 but do not fall in Fig. 3 C. It means that pure science was more seriously interrupted by war than was engineering. All indicators fall in the interval of 1945–49 in Fig. 3 C but some fall in the same time in Fig. 3 B.

There would be many reasons to explain the phenomena but we wish to think the phenomena rather depend upon the selections of sample population of Taikei itself. We prefer to notice the rising of indicators in Fig. 3 C since 1950 whereas those in Fig. 3 B are still in through.

### 3.3 Influence of unusual events on the development of sciences

As told in 2.2, we attached importance to an old viewpoint of classification. Chinese and Korean historians devoted several chapters of a history of a dynasty to recording wind calamities, floods, earthquakes, blights, solar eclipses, comets and so on, in a title of "five elements" as the heaven's prognostication of change of the era. Some of those events or phenomena had serious influences on developments of societies. We selected unpredicted and inevitable events or phenomena as unusual and catastrophic events or phenomena, and made up the seventh bough of classification of Key-words of the thesaurus. Distribution of Key-words of documents concerning "Unusual Events, Catastrophes, or Phenomena" is shown in Fig. 3 D. We find that these unusual events have significant relationship to developments of science and technology.

Among unusual events, catastrophes or phenomena, the most important in Japan were wars and earthquakes. We also took out a set of Key-words of military affairs and ordinances and a set of those earthquake disasters, and obtained graphs of them. Fig. 3 E shows the distributions of those concerning military affairs and ordinances. We can easily find some close relationship between wars and military sciences. The modern military sciences arose about 1840 after the shock of hearing of the Opium War. (In this graph, all indicators rise in 1850 because of the reason which we have explained before.) During 1840 to 1860, Japanese learned military sciences through the Dutch language i. e. "Rangaku". The Meiji Restoration (1868) and Seinan Civil War (1877) accelerated military sciences and Key-words of *policy and economy* became very high. The Russo-Japanese War and World War I rather depressed the military sciences but the Sino-Japanese War (of 1930) and World War II re-activated them. After World War II, Key-words of *policy and economy* and *social and labour problems* rose very high. Both of them indicate the social mess under the occupation of United States military forces and indicate the guilty feeling of the scientists to military sciences. In March 1954, the crew of the fishing boat, Fukuryu-maru No. 5, was fatally exposed to radio-activity resulting from U. S. H-Bomb tests and Japanese experienced a great panic concerning radio-active fallout. Both indicators still remain high in 1950–54.

Fig. 3 F shows the distributions of those concerning earthquake disasters.

One of the greatest earthquakes of the "Ansei" ere occured in Tokyo in 1855, and the first peak coincides with this event.

After the Meiji Restoration, foreigners surveyed Japanese land and during this time there occured the Yokohama Earthquake in 1880. Professor John Milne of Tokyo University established the Earthquake Association of Japan as the first society of its kind in the world, and many professors from Europe also showed much interest in earthquakes in Japan. The indicator of *science and research* rose in the interval 1870–74 with the land surveying by foreigners and sharply rose after the Yokohama earthquake. It is necessary to point out that the contributions of foreigners to earthquake disasters were higher than those of Japanese in the period to 1884. Japanese scholars, who were taught by foreign professors, also showed interest in earthquake phenomena and when the Nobi earthquake struck Nagoya in 1891, they investi-

gated all surrounding districts. Almost all of the brick buildings (which Prof. Milne had proposed as rigid earthquake-proof structures, in preference to the traditional wood and stone buildings) were destroyed by the earthquake.

There arose re-investigation of constructing earthquake-resistant and fire-proof buildings, therefore the indicator of *engineering and industry* rose high in the interval 1890–94. Investigations to prevent earthquake disasters continued except during the period of the Russo-Japanese War, but they were fruitless in the fields of improving architecture, of town-planning or of other real branches of engineering. Indicators of *"lives"* (including town), and *social and labour problems* were as high as those of *science and research, engineering and industry*, and *policy and economy* when the great Kanto earthquake struck Tokyo and its surrounding industrial zones in Sept. 1923.

As great earthquakes in populous areas occur infrequently outside Japan, foreigners had little knowledge of such phenomena or methods of construction of earthquake-proof buildings. Unlike foreigners, most Japanese were accustomed to earthquakes, therefore, they were less interested in earthquakes. In addition they could have hardly learnt about earthquake or earthquake-proof buildings through Dutch languages. Thus foreigners from Europe or America had to be the main contributors to earthquake disasters at the beginning. Numbers of foreigners as Key-word disappear since 1900. We are able to give several localized fields of sciences as examples in comparison with earthquake disasters which have mainly foreign contributors at the beginning. We mention a case of botanical nomination in Linnaean Classification in Japan. As an example K. P. Thunberg (1743?–1828), who was in Japan from 1775 to 1777, investigated plants and published "Flora Japonica". In Makino's pictorial book of Japanese Flora (4), we find 107 of the total 2322 botanical species also named by Thunberg.

Amongst Thunberg's nominations, we find *Iris japonica* (purplish white flower in early May), *Lilium longiflorum* (white lily), *Camellia sasanqua* (small camellia in November to December), *Nendina domestica* (fruits as of Christmas holly are used for ornaments of New Year) and so on. Next came P. F. Siebold to Japan, from 1823 to 1829 and 1859 to 1862 who nominated 57 species by himself and 171 species with D. G. J. Zuccarini (5). We find 120 species nominated by Siebold or Siebold et Zuccarini in the said Makino's pictorical book, among them are *Prunus donasium* Sieb. (popular cherry), *Magnolia parviflora* Sieb. et Zucc. (white noble flower with a boss of climson stamens in the centre, for tea ceremony) and many trees as *Pinus parviflora*. After Siebold, F. A. W. Miquel gave names to about 70 species of plants.

An early Japanese godfather was Tomitaro Makino and we count 249 species nominated by him in the said pictorial book. Plotting numbers of botanical species nominated both by foreigners and by Japanese, we have the same tendency as that for the contributions of persons to earthquake disasters, namely that a line of numbers by foreigners has a peak at the beginning and a line by Japanese at a later time.

### 3.4 Contributions by foreigners

The distributions of foreigner's names are shown in Fig. 3 G. In this figure, we reach the peak of Key-words in *science and research* in the interval of 1880–1884, and the peak in *engineering and industry* in the interval 1890–1894. On the other-hand, the peak in *policy and economy* is in the 1870–1874 period, and the peak in *lives, medical treatments, and education* is the same period, and so on. These tendencies were considered to be the effects of the Meiji Government's policy to develop science and engineering by assistance from employed foreign scholars and experts. This policy is well known by the name of "Fukoku-kyohei" i. e. the promotion of national industries for wealth and armament.

Thus we analyse these reaons for appearence of many foreign scientists and engineerers in that time, depending upon the government of promotion of national industries, as shown below:

i) Engineering field: including ship-building, civil engineering, spinning industry and steel & military industries.
ii) Land survey field: including mineral resources, land measuring to make modern maps, and meteorological observations.
iii) Education field: including establishment of universities, learned societies in science and engineering. etc.

About 1890, the government policy of employing Europeans in engineering fields was ended, then the visiting professors of universities or the visits of famous foreign scientists appear in Key-words. Thus the appearence of foreign names until the mid-Meiji era and the succeeding periods is due to different reasons.

i) Pure science fields such as chemistry and physics, are not so depressed in frequency as the local science field: geology, botanical-geography and so on, which rapidly decreased in frequency after 1900.
ii) In general, engineering and industry and social and labour problems gradually decreased.
iii) Thought, philosophy, religion, fine arts, reached a peak of distribution in the period 1935–39.

Finally, the distribution of frequency of Japanese names in appearance of Key-words is herein omitted.

### 3.5 Policy-law-institute relationship

Finally, we present some interesting graphs. We selected five themes, *"Objects of Policy"* (50 Key-words were selected), *"Administrative Organization"* (85 Key-words), *"Code"* (7 Key-words, as all indexing numbers of documents for 51 acts, laws and rules were duplicated in some broader terms), *"Universities"* (42 Key-words) and *Research Institutes* (36 Key-words) and had the same process as former ones. When we obtained the data, we gathered indicators of *science and research* of the five themes and made up a graph as shown in Fig. 4 A. In this graph, we can see some relationships between the five themes in the field of science and engineering. Similarly, we plotted the indicators of *engineering and industry* in Fig. 4 B, and *lives, medicines and education* in Fig. 4 C. We also plotted the indicator for the theme *"Companies"* (72 important companies and Key-words of "Company") in Fig. 4 B.

As a rule, policies come after some events or problems occur. Policies demand the making of laws, bureaus make suitable changes in their system to solve the problems, and some contributions are made by research institutes and new institutes may be established. Universities would also make some contributions to them. In our work, as said in 3.1 (1), we could hardly select minor events. We merely analysed some macroscopic relationships between the 5 themes.

In Fig. 4 A, *policies* for science grew following the Opium War, *administrative organizations* were gradually consolidated (and firmly established) during 1870 to 1884. Universities were settled, legal systems and institutes were also completed. These tendencies were further promoted after the Russo-Japanese War and reached a peak after 1910. Similar phenomena were found before World War II. In 1945 we had a lot of prohibitions and withdrawels of military sciences including, nuclear engineering and aero-engineering.

By the way, *administrative organization* followed *policies* with a delay of a five-year interval but *policies, institutes and legal systems* occur simultaneously in five-year intervals in the field of *science and research*.

In the field of *Engineering and industry* including military ordinances, in Fig. 4 B, policies, legal systems and bureaucratic organization rose following the Opium War. In this figure, we must pay attention to the fact that the indicator for *code* follows that for *objects of policies* with a five year delay and the fact that the indicator of *research institute* follows that for *objects of policies* with a 10 year delay. We think that one of the major reasons is that research institutes in the field of *engineering and industry* are established not voluntarily nor spontaneously but due to great necessity by governments and the economic world. These tendencies are also sharply distinguished between research institutes for major science and those for minor sciences.

Similarly to Fig. 4 A, research institutes grew since 1910 to promote domestic industry, as did influences of the government i. e. the indicator of *administrative organization* grew with time. *Universities* made few contributions to *engineering and industry* in Japan during the last one hundred years. In Fig. 4 C, we show the indicators of *lives, medicines and education* (including town), however, these indicate only the tendencies of lives (town), and medicine as we omitted three volumes for education.

The most important problem after the Meiji Restoration was the organization of army surgeons and the change over from traditional Chinese medicines to European ones which were suitable for the army in the field of surgical operations. New universities were quickly established and existing ones re-organised to provide doctors. Through one hundred years, the indicator of *research institutes* also does well, but that of *universities* corresponds less to the *objects for policies*. As the final figures provide very complicated problems for analysis, we must make more considerations to explain these phenomena. These three figures are presented not as results for our analysis but as some information for the reader.

## References


1) Tomita, T., and Hattori, K.: History of Science Society of Japan (Eds.) *Nihon Kagaku Gijutsu-shi Taikei* (History of Science and Technology in Japan), 25 vols., 1964–1970. In: Japanese Studies in the History of Science (1970) No. 9.

2) – : Compilation of a thesaurus and total index for *Nihon Kagaku-Gijutsu-shi Taikei* by means of a computer. *ibid* (1972) No. 11.

3) – : Outline of a thesaurus of *Nihon Kagaku-Gijutsu-shi Taikei* with Heading List of Classified Key-words. *ibid* (1973) No. 12.

4) Makino's pictorial Book of Japanese Flora, New Edition for Students. Hokuryukan 1967.

5) Tei-ishi Ishi-yama: Siebold's Work on the recording of Botany in Japan. In: Japanese J. Hist. of Biol. (February, 1974) No. 25.

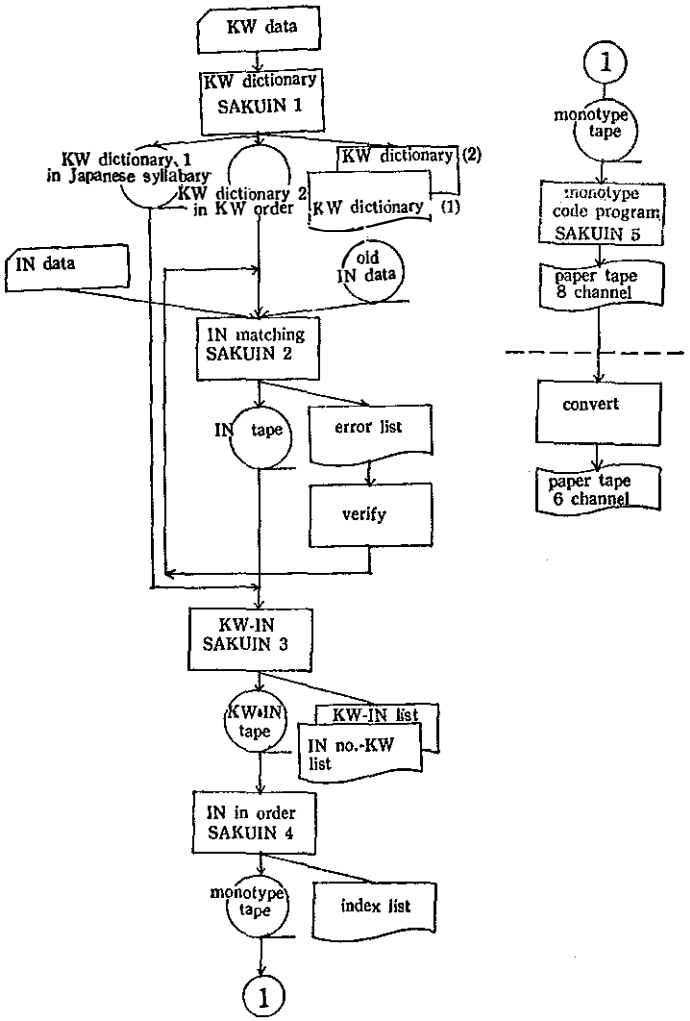




*Fig. 1 The general flow chart*

*Table 1: List of the titles, number of documents, and numbers of words in the indices*

| Volume | Title | No. of pages* | No. of chapters | No. of documents | No. of words in index ($c = d + f$) | No. of personal names ($d$) | No. of foreign names ($f$) | $c/b$ | $c/d$ | $f/d$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Outline history I | 582 | 14 | 257 | 829 | 316 | 29 | 3.24 | 0.381 | 0.092 |
| 2 | Outline history II | 596 | 15 | 217 | 887 | 329 | 53 | 4.09 | 0.370 | 0.161 |
| 3 | Outline history III | 550 | 11 | 213 | 1,359 | 301 | 20 | 6.37 | 0.222 | 0.065 |
| 4 | Outline history IV | 532 | 10 | 229 | 1,179 | 116 | – | 5.14 | 0.098 | 0 |
| 5 | Outline history V | 558 | 13 | 218 | 1,062 | 119 | 19 | 4.87 | 0.112 | 0.160 |
| 6 | Philosophy | 552 | 10 | 187 | 758 | 293 | 43 | 4.05 | 0.386 | 0.150 |
| 7 | International | 566 | 14 | 152 | 1,794 | 558 | 464 | 11.82 | 0.311 | 0.800 |
| 8 | Education I | 562 | 14 | 259 | 1,405 | 407 | 94 | 5.42 | 0.290 | 0.231 |
| 9 | Education II | 542 | 16 | 205 | 1,521 | 509 | 58 | 7.41 | 0.335 | 0.114 |
| 10 | Education III | 578 | 14 | 207 | 1,779 | 270 | 20 | 8.60 | 0.152 | 0.739 |
| 11 | Natural environment | 594 | 17 | 229 | 1,834 | 681 | 41 | 8.00 | 0.371 | 0.601 |
| 12 | Mathematical science | 598 | 16 | 263 | 1,385 | 407 | 85 | 5.26 | 0.294 | 0.208 |
| 13 | Physical Sciences | 576 | 12 | 198 | 1,588 | 453 | 79 | 8.00 | 0.286 | 0.174 |
| 14 | Astronomy & earth science | 614 | 15 | 349 | 1,524 | 612 | 87 | 4.37 | 0.402 | 0.142 |
| 15 | Biological Science | 538 | 10 | 194 | 1,073 | 404 | 121 | 5.49 | 0.376 | 0.300 |
| 16 | Civil engineering | 542 | 10 | 173 | 1,377 | 207 | 27 | 7.95 | 0.151 | 0.130 |
| 17 | Architecture | 590 | 12 | 275 | 533 | 199 | 50 | 1.94 | 0.374 | 0.251 |
| 18 | Mechanical engineering | 580 | 19 | 208 | 1,380 | 296 | 36 | 6.63 | 0.214 | 0.123 |
| 19 | Electrical engineering | 544 | 14 | 240 | 1,110 | 275 | 51 | 4.63 | 0.248 | 0.185 |
| 20 | Mining & metallurgy | 554 | 12 | 211 | 1,392 | 493 | 71 | 6.59 | 0.354 | 0.144 |
| 21 | Chemical engineering | 542 | 10 | 243 | 1,051 | 481 | 16 | 4.33 | 0.457 | 0.033 |
| 22 | Agriculture I | 572 | 11 | 231 | 1,783 | 251 | 22 | 7.71 | 0.141 | 0.087 |
| 23 | Agriculture II | 536 | 10 | 149 | 1,544 | 165 | 10 | 10.37 | 0.107 | 0.061 |
| 24 | Medicine I | 540 | 10 | 153 | 1,013 | 291 | 61 | 6.67 | 0.254 | 0.210 |
| 25 | Medicine II | 570 | 9 | 190 | 1,083 | 222 | 27 | 7.86 | 0.205 | 0.122 |
| | Total | 14,088 | 318 | 5,450 | 32,243 | 8,665 | 1,584 | | | |
| | Average | 564 | 12.7 | 218 | 1.290 | 347 | 63.4 | 5.92 | 0.269 | 0.183 |

* Number of pages does not include preface pages etc. because irrelevant.

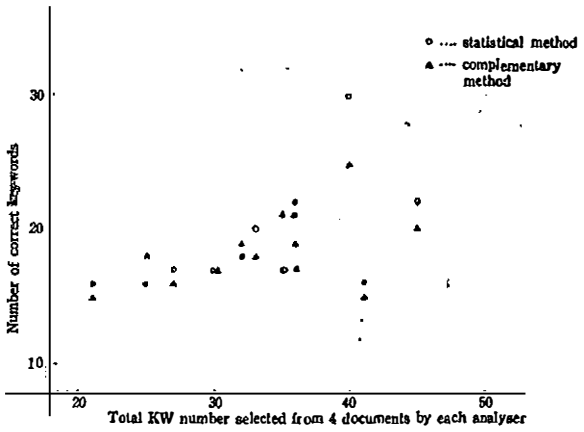
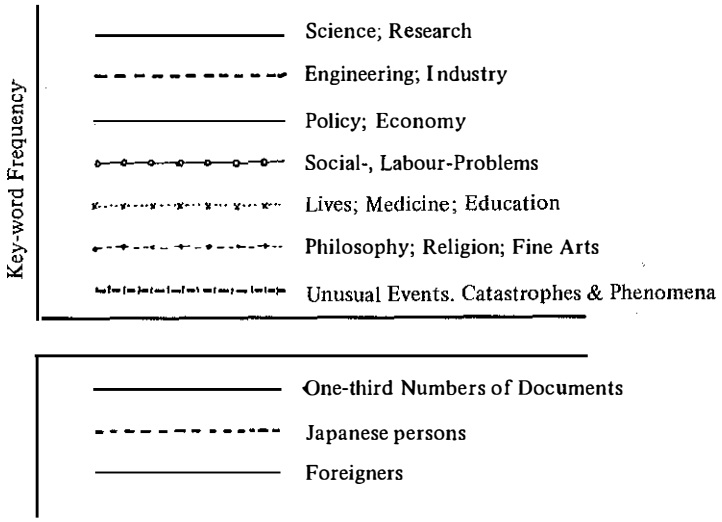



*Fig. 2 Total correlation diagram of KW number and number of correct KW selected by each analyser from 4 documents*




*Legend for Fig. 3A–3G*

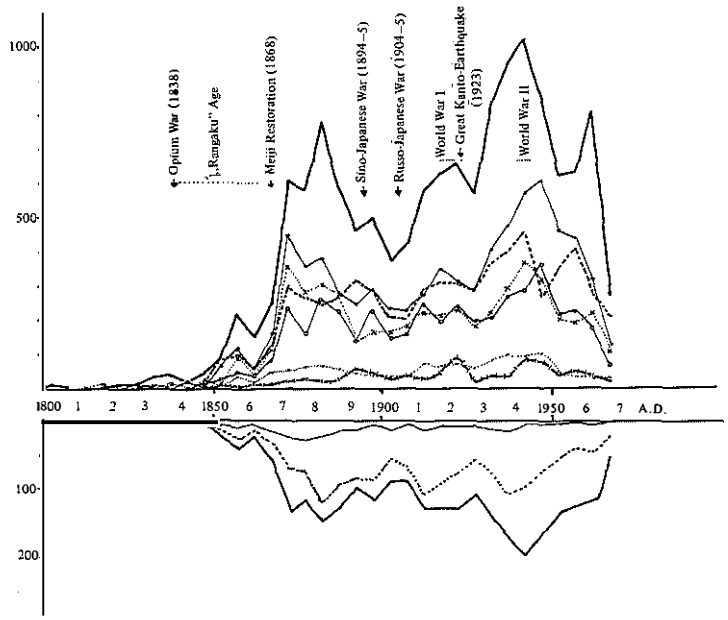



*Fig. 3A All documents in 21 volumes*

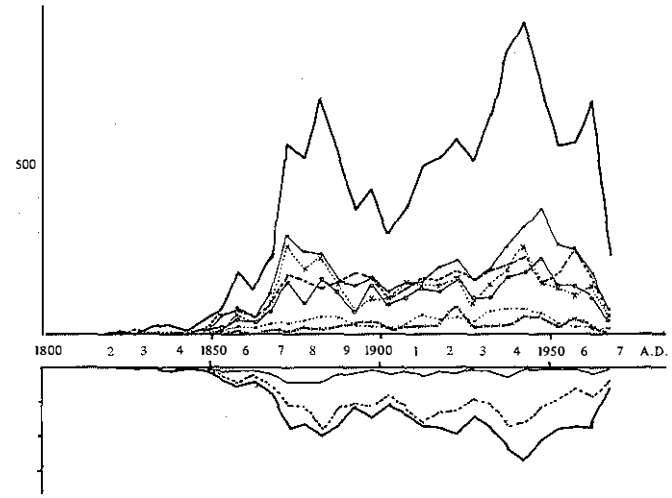



*Fig. 3B Science and research*

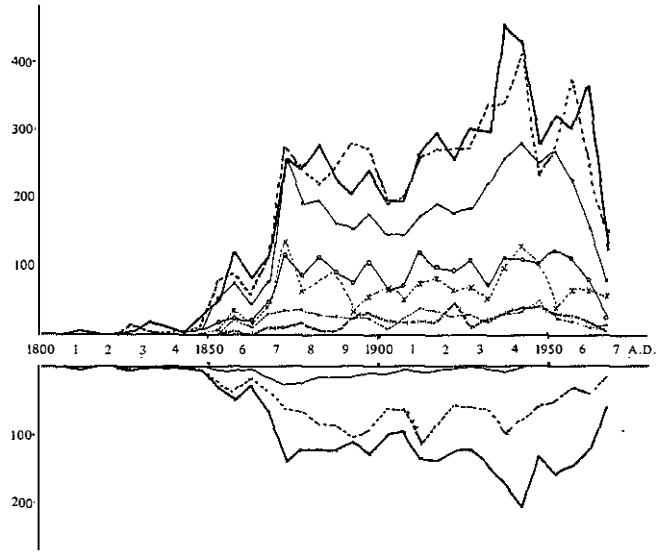



*Fig. 3C Engineering and industry*

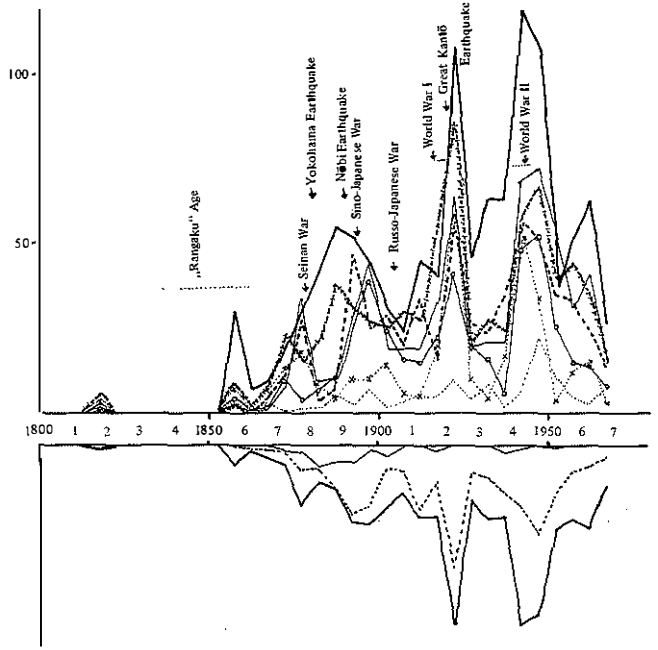



*Fig. 3D Unusual events, catastrophes and phenomena*

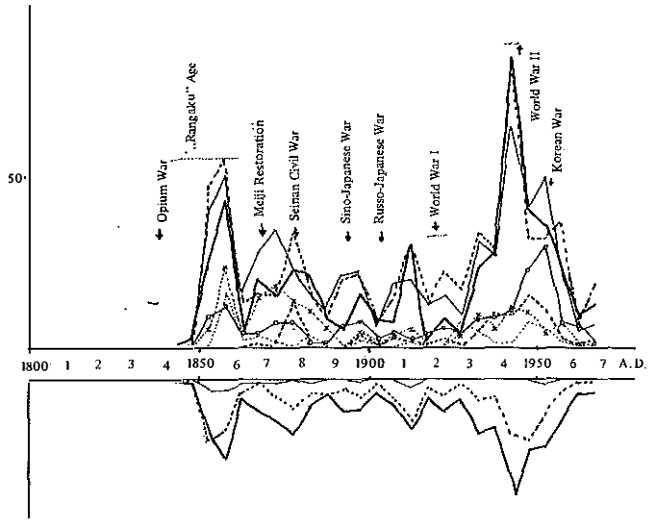



*Fig. 3E Military affairs and ordinances*

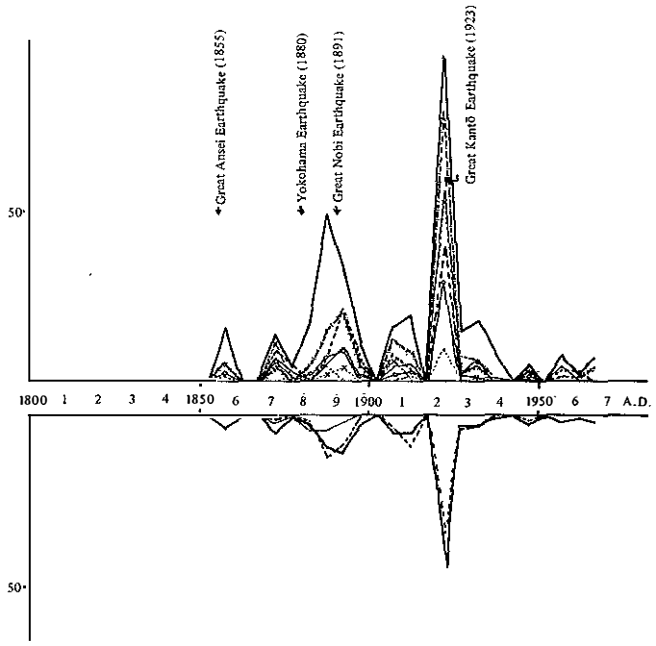



*Fig. 3F Earthquake disasters*

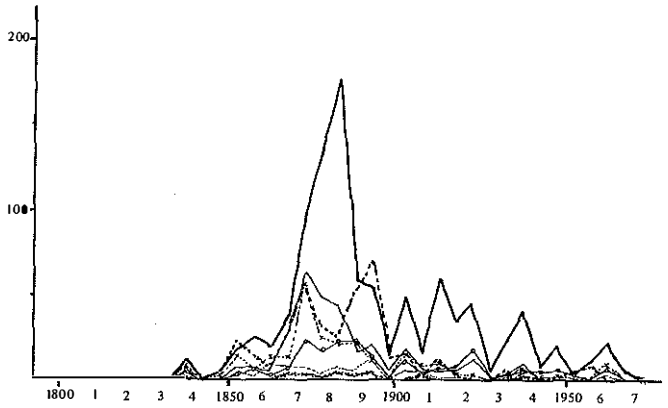
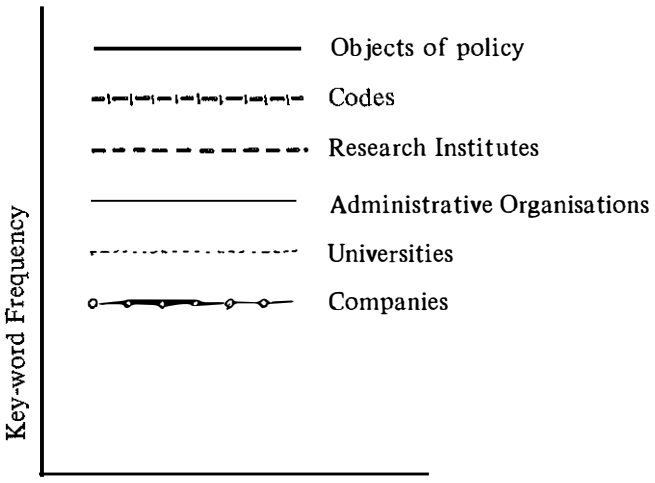



*Fig. 3G Contribution of Foreigners*




*Legend for Fig. 4A–4C*

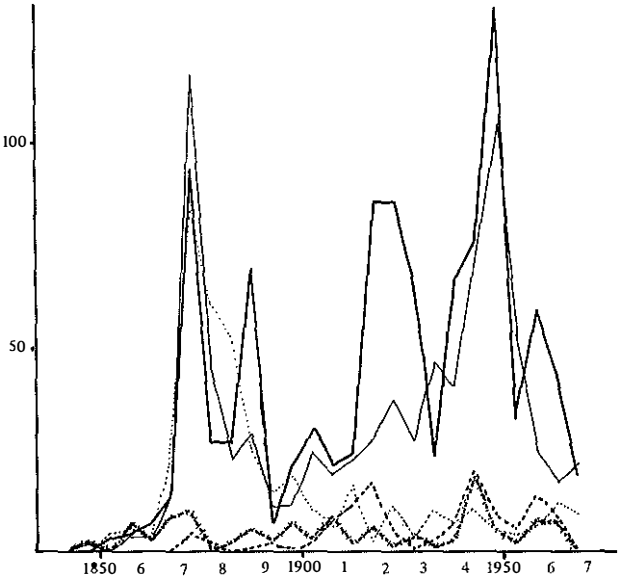



*Fig. 4B Engineering and industry*


*Fig. 4A Science and research*


*Fig. 4C Lives, medicine, and education*