

Foreign Literature Challenges to Chemical Abstracts Service[†]

JOHN T. DICKMAN and GERARD O. PLATAU*

Chemical Abstracts Service, Columbus, Ohio 43210

Received July 6, 1982

Nearly 75% of the journal literature abstracted by *Chemical Abstracts* (CA) originates from some 140 countries other than the United States and is written in 55 different languages. In addition to the 400 000 papers abstracted each year, CA covers by abstract or citation some 170 000 patents yearly, 150 000 of which are from 25 foreign countries and 2 industrial-property organizations. Editorial processing, both bibliographic and technical, is handled by staff who are experts in their specialized areas of scientific interest and who are also proficient in the language of the original document. Problems dealing with acquisition, transliteration, abstracting, indexing, and document delivery are identified and discussed.

INTRODUCTION

Chemical Abstracts (CA) came into being as the result of several years of effort to bring worldwide recognition to American chemists whose accomplishments seemingly were being overshadowed by the great European scientific tradition. To be sure, before the turn of the century chemistry in America was well established and flourishing. Yet American chemists were lacking in deserved international recognition. They possessed no English-language abstracting and indexing service, having to rely instead on the German *Chemisches Zentralblatt*. In 1895 the progenitor of CA, *Reviews of American Chemical Research*, first appeared as a supplement to *Technology Quarterly*, published by the Massachusetts Institute of Technology, and 2 years later became part of the *Journal of the American Chemical Society*. It sought to rectify this deficiency by printing English-language abstracts of chemical research, initially restricting its coverage to authors residing and publishing in the United States. In 1902 its coverage was extended to include research reported by American chemists in foreign publications. Five more years of evolutionary development led to the Jan 1, 1907, birth of CA as a free-standing publication of the American Chemical Society, covering not only U.S. research but new chemistry from around the world. It is interesting to note that the first two abstracts published in Vol. 1, Issue 1, cited papers from the German *Zeitschrift fuer Angewandte Chemie* by German authors. CA has continued for 75 years to serve as the most comprehensive, English-language source of worldwide chemistry and chemical engineering research information.

DIMENSIONS AND DEFINITIONS

Coverage of the foreign scientific and technological literature of chemistry indeed presents a formidable challenge to both the technical expertise and to the man-machine editorial processing interactions that together constitute Chemical Abstracts Service (CAS). Chemical literature emanates from the developing countries of the world as well as from the nations that lead in advanced research and development. On the basis of 1981 statistics, the Soviet Union led the foreign contributors by providing 16.6% of the total chemical journal literature abstracted by CA (Figure 1). A total of 140 countries, in addition to the U.S., added to the body of chemical knowledge through the scientific literature.

Although the Soviet lead has persisted since Sputnik literally rocketed the world into the space age in the late 1950s, there has been a dramatic reduction in the percentage of Russian contribution to the foreign literature over the past 10 years. Whereas that percentage stood at 25% in 1972, it has dwindled

to 16.6% in 1981. The cause of this reduction is speculative, but CAS Director Baker¹ has suggested that the concepts of strong control and reduction of expenditures, both of which have been identified by the Soviets as basic to their implementation of an automated information system, may indeed adversely affect free flow of information beyond national borders. These effects may already be making their appearance in these statistics.

From Figure 2 it is apparent that English is the worldwide language of chemistry, with Russian, Japanese, German, and French leading the foreign-language contributors. The total list of languages other than English represented by CA abstracts in 1981 is given in Figure 3. CA also covers the chemical patent literature consisting of granted patents and examined and unexamined patent applications from 25 foreign countries and two international organizations (Figure 4) and published in 16 different languages.

In addition to the dimensions of the foreign literature, a few definitions are in order. Serial publications, exemplified by the familiar scientific periodical with its articles describing original research, make up the bulk of the scientific literature. Over 10 000 serial publications are regularly examined for documents whose subject matter is suitable for coverage in CA. A document is defined as any abstractable item such as a journal article, a review, or technical report. A document package refers to a collection of documents such as an issue of a journal or a conference proceedings; 65 000 such packages are reviewed each year.

LANGUAGE SKILLS AND TRANSLITERATION

It is apparent that scientific expertise represents only part of the skill which technical staff must display in dealing with these raw materials in the manufacture of *Chemical Abstracts*. Without the appropriate language skills to accompany knowledge of the subject matter, document-package evaluation, selection of documents, and the generation of English-language abstracts and index entries could not be accomplished. While translation into English is a skill required at many stages of the operation, translators, as such, are not employed by CAS. There is no large-scale translation of complete foreign-language documents. Rather, bilingual comprehension is a necessary tool employed throughout the editorial process from the evaluator of the literature to the document analysts and other editorial staff who produce the bibliographic citations, the abstracts, and index entries.

An operation related to translation, yet distinctly different from it, is transliteration. It is necessary to convert the eastern European, Middle Eastern, and Oriental languages, which do not use the modern Anglo-European alphabet, into readable roman characters. This process is called transliteration and traditionally is based on sets of specifically defined character substitutions designed to duplicate the sound represented by

[†] Presented on March 29, 1982, as part of the symposium on Problems with Foreign Literature during the 183rd National Meeting of the American Chemical Society, Las Vegas, NV.

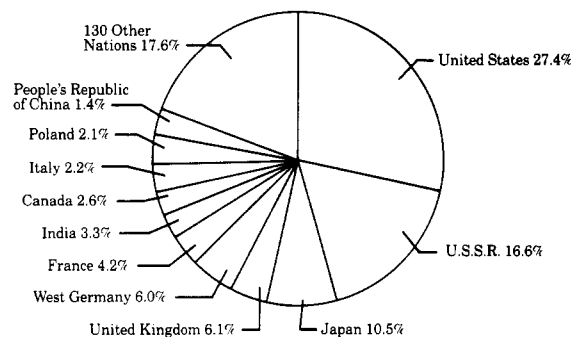


Figure 1. Sources of the world's chemical journal literature in 1981.

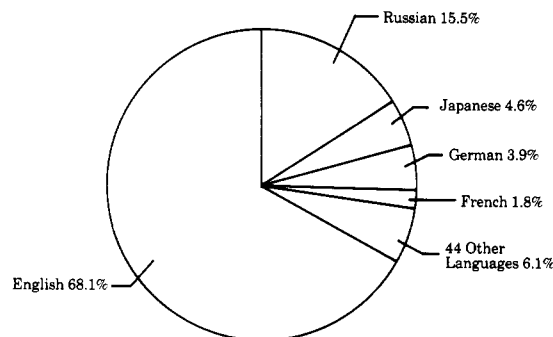


Figure 2. Languages of the world's chemical journal literature in 1981.

AFRIKAANS	GREEK	PUNJABI
ALBANIAN	GEORGIAN	PERSIAN
ARABIC	GERMAN	POLISH
ARMENIAN	HEBREW	PORTUGUESE
AZERBAIJANI	HINDI	ROMANIAN
BASQUE	HUNGARIAN	RUSSIAN
BELORUSSIAN	INDONESIAN	SERBIAN
BULGARIAN	ITALIAN	SERBO-CROATIAN
CAMBODIAN	JAPANESE	SLOVAK
CHINESE	KOREAN	SLOVENIAN
CROATIAN	LATVIAN	SPANISH
CZECH	LITHUANIAN	SWEDISH
DANISH	MACEDONIAN	THAI
ESTONIAN	MALAYAN	TURKISH
FINNISH	NETHERLANDISH	UKRAINIAN
FRENCH	NORWEGIAN	VIETNAMESE

Figure 3. Languages of the world's chemical journal literature abstracted by CAS in 1981.

AUSTRIA	GERMAN	POLAND
AUSTRALIA	DEMOCRATIC	ROMANIA
BELGIUM	REPUBLIC	SOUTH AFRICA
BRAZIL	GERMANY,	SPAIN
CANADA	FEDERAL	SWEDEN
CZECHOSLOVAKIA	REPUBLIC OF	SWITZERLAND
DENMARK	HUNGARY	UNION OF SOVIET
FINLAND	INDIA	SOCIALIST
FRANCE	ISRAEL	REPUBLICS
	JAPAN	UNITED KINGDOM
	NETHERLANDS	UNITED STATES OF AMERICA
	NORWAY	
EUROPEAN PATENT	WORLD INTELLECTUAL	
ORGANIZATION	PROPERTY ORGANIZATION	

Figure 4. CA patent coverage by country in 1981.

the characters in the original language. Problems arise as a result of the absence of one internationally accepted standard for many of these languages that require transliteration. Obviously, it depends upon who is trying to duplicate a particular sound. A German, for example, might insist on a rather guttural rendering of a certain Arabic symbol, while an English linguist would pronounce the same symbol with a much softer sound and give it a different Latin character accordingly.

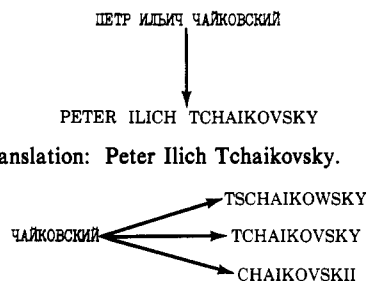


Figure 5. Translation: Peter Ilich Tchaikovsky.

Figure 6. Three-way translation: Tschaikowsky, Tchaikovsky, Chaikovskii.

While the series of Cyrillic characters at the top of Figure 5 is probably only immediately interpretable by one familiar with the Russian language, when transliterated the name of Peter Il'ich Tchaikovsky is universally recognized by classical music enthusiasts throughout the world. However, if this famed composer had also been a chemist, as was his contemporary and compatriot Alexander Borodin, his name would appear in the CA Author Index as shown at the bottom of Figure 6. It would begin with C and be alphabetized accordingly. The transliteration scheme followed by CA for Russian and other Cyrillic-character languages is essentially that promulgated by the American National Standards Institute (ANSI).² The more familiar form of the name is the center version and is the one generally used on record album covers and in concert programs. It is a somewhat Anglicized version of an older Germanic system shown at the top of Figure 6, both of which were operative long before the current ANSI standard was issued. The continued popular use of the familiar center version is the result of tradition in the musical world and in no way indicated any preference for the system in other applications. Indeed, the ANSI standard is by far the most widely used Russian transliteration scheme in science and industry today, and it also addresses the other five major slavic languages: Ukrainian, Belorussian, Serbian, Macedonian, and Bulgarian. Yet, truly internationally accepted transliteration standards, which deemphasize attempts to duplicate sounds and focus instead on deriving a universally accepted set of symbols, are still needed for effective interlingual communication.

As the People's Republic of China emerges from decades of scientific and cultural reclusion, its contribution to the chemical literature of the world is increasing significantly. For many years, the Wade-Giles system of transliteration, named for its originators, two 19th Century British diplomats, Thomas Wade and Herbert Giles, was used to convert the Chinese ideograms to English syllables that as closely as possible approximated, at least to the ears of the inventors, the singsong quality of the spoken Chinese language. In Jan 1979, the government of the People's Republic of China announced that henceforth all official business would be conducted in a new system, the "Pinyin" system,^{3,4} for transcribing Chinese names and places into the Roman alphabet. Although conversion has been a slow process, today nearly all titles of scientific publications originating in the People's Republic of China bear a transliterated title in the Pinyin system. The Chinese community outside the People's Republic of China, for example, Taiwan and Hong Kong, has not accepted the conversion and is not likely to do so in the foreseeable future. Multiple transliteration problems of this nature obviously impact upon industry, government, and academia, and this area is under extensive investigation at the present time. CAS has established a procedure to ensure bibliographic integrity and consistency. CAS currently identifies in CA the Chinese-character citation in Figure 7 from the *Chinese Pharmaceutical Bulletin* according to the lower name, that of the new Pinyin system. The Wade-Giles rendering of these ideograms, and the name

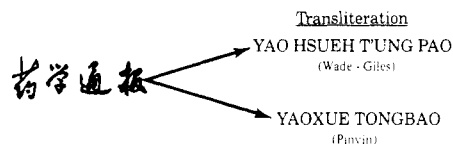


Figure 7. Translation: Chinese Pharmaceutical Bulletin.

by which this journal has been identified in CA since the 1950's, is illustrated by the upper name. This kind of dual nomenclature is identified for the user in the CAS Source Index by means of a cross-reference. Regardless of which system ultimately gains worldwide acceptance and to assist the searcher in retrospective searching, a cross-reference will always interlink the two names in the Source Index.

The handling of the names of Chinese authors and geographic locations is influenced by this dualistic transliteration problem. Often a partial listing of Pinyin-transliterated author names appears in English-language tables of contents in journals from the People's Republic. Such listings in journals from Taiwan, Hong Kong, and elsewhere outside the People's Republic appear in the Wade-Giles version. In both cases, in order to complete the author indexing process, CAS must transliterate any remaining names from the papers themselves by using the transliteration scheme that fits the source. Thus far, no problems to CAS or its users have resulted, simply because the two factions do not publish in each others' journals. Should that situation ever change, a single Chinese transliteration standard would have to be given high priority by the standards writers. Chinese cities and provinces bear Pinyin-transliterated names in CA, except for a few deeply ingrained traditional names like Canton and Chungking, which predate even the Wade-Giles system. Interestingly enough, Shanghai is transliterated as such by either system.

Transliteration is no less a problem with the Japanese literature, for which there are extant at least three systems of character romanization. The system used almost universally today, in Japan and abroad, and that which forms the basis for the current ANSI standard⁵ is a modified Hepburn system, named for an American Presbyterian missionary, James Hepburn, who originated it in the late 1800s.

DOCUMENT ANALYSIS

While transliteration provides a solution to the problem of romanizing names of publications, authors, and places, the foreign language itself, whether written in roman characters or not, can present a significant challenge to the scientist in need of information. It is here that CAS provides what is perhaps its most important service—the link between the creator of information and its user, when the two communicate in different languages. The importance of the abstract, in contrast to an article title or an author name alone as the link has been amply and ably described by Russell J. Rowlett, former Editor of *Chemical Abstracts*. In his 1981 Miles Conrad Memorial Lecture,⁶ he envisions a continued need for the abstract well into the foreseeable future simply because of its invaluable role as an information-packed filter to aid in determining topic relevance in retrospective searching. When this role serves as a key to unlock the foreign-language barrier, chemical information becomes more accessible to all.

Foreign-language proficiency for a major portion of the document-analysis staff at CAS is second in importance only to scientific and technical competence in the subject areas in which they work. Of the nearly 300 document analysts who generate the vast majority of the 9000 abstracts and associated index entries produced each week, 60% are capable of performing their task from at least one of the 55 languages other than English in which the scientific literature is written. As was mentioned earlier, abstracting does not involve a verbatim

written translation of a complete document into English prior to preparation of the English-language abstract. Rather, the document analyst reads the document in the language in which it is written and simultaneously creates the English-language abstract in something of a combination "comprehend and condense" operation.

The document analyst is a relatively new breed of chemical documentalist. From the beginnings of CA and continuing for the first half century of its existence, abstracts were produced by a dedicated core of volunteer abstractors located around the world. Chemists, chemical engineers, and scientists of all fields whose interests extended to chemistry abstracted the world's scientific literature as an avocation. The small honorarium they received was not the motivation for their efforts. Rather, they volunteered their time and their expertise because abstracting provided an opportunity for them to read the current literature, to exercise proficiency in a foreign language, or, quite altruistically, to make a small and personal contribution to effective communication among scientists.

At the height of the volunteer abstractor program in the mid 1960s nearly 3300 volunteers prepared abstracts for CA. Editorial staff functions primarily consisted of editing to assure accuracy and format consistency and, of course, the monumental indexing task. But the volunteers produced the abstracts. Over half of them resided outside the United States. Their numbers provided the language capability necessary to do the job.

Despite the heroic efforts of this corps of abstractors, exponentially growing chemical research was threatening the very existence of the medium that was attempting to document it. The CA production system that worked so well for so many years could no longer handle the worldwide onslaught of scientific and technical information. It was to be replaced by one that would not only keep up with the proliferation of information but would, in fact, vastly improve timely access to and delivery of chemical information to those who needed it. The core of this current system is the computer and its technology fed by the intellectual input of the document analyst, who is at once scientist/linguist/abstractor/indexer/editor.

The volunteer ranks have shrunk dramatically over the past 15 years. The in-house document analysis staff has gradually but steadily assumed the input to the CAS computer-based information processing system. A mere handful of volunteers remains to serve mainly in urgent or emergency situations. The rarity of such a talent combination as (to cite a hypothetical example) a spectral chemist who reads Turkish seems, for the foreseeable future, to mandate reliance on the qualified volunteer to handle the few, but nonetheless potentially important, documents that might address that topic. It is an extremely difficult task to acquire in-house specialists for every conceivable subject discipline and language, although this is the ultimate goal.

Additionally, over the years the volunteer abstractor program has developed certain teams within a team, so to speak, in several countries important to the chemical world. Abstractors in Japan, Poland, and Czechoslovakia comprise such teams whose mission it is to provide abstracts from specified journals and patents published in their particular country. Team leaders reside in those countries to make the appropriate assignments to their team members and to monitor CAS coverage of a large portion of the literature emanating from that country, thus relieving the in-house staff of some of the pressures of assuring completeness of coverage.

FOREIGN PATENT LITERATURE

While the scientific periodical, the technical report, and the conference and symposium proceedings represent the bulk of

the scientific and technological documentation of the world, one important segment, the patent literature, must not be overlooked. The technology by which mankind has advanced during the past couple of centuries is largely documented in the patent literature of the world.⁷

Patents have been covered in CA since Issue No. 1 in 1907. In fact, over 32% of the 11 847 abstracts printed in CA's first year of existence were patent abstracts, largely U.S. documents but including patents from the United Kingdom, France, and Germany. A brief statement of patent coverage policy in Issue No. 1 of Vol. 1 of CA begins with the uncompromising statement "All civilized countries of the present time have some kind of Patent system, usually more or less modeled after the Patent system of the United States." This sweeping statement from early CA is probably not an overstatement. Actually over 100 countries grant some sort of patent protection. To obtain optimal coverage of the world's patent literature with full consideration given to economic, accessibility, availability, and timeliness factors, CAS covers the patent documentation from 26 countries and 2 international organizations. This extensive coverage yields over 170 000 patent citations annually, both as abstracts and as patent equivalents cited in CA's Patent Index.

Over the years the procurement of foreign patents has proven more of a problem than the language. The 25 foreign countries currently covered have not constituted a fixed list. Country coverage has changed from time to time as the result of document availability and logistics considerations. Direct contact with most foreign patent offices throughout the world assures a continuous and timely flow of patent documents for abstracting. However, in some countries (for example the Republic of South Africa), it is necessary to work through acquisition agents because the patent documents cannot be acquired directly from the patent office. Soviet-bloc countries are especially slow in filling patent orders. Czechoslovakian patents are not available for some 2 years after they are announced to the world. Citation of Russian patents is expedited by using the abstracts published in the Russian counterpart of the U.S. Patent Gazette as a basis for the preparation of the CA abstract. While the quality of these Russian patent summaries in terms of the amount, type, and accuracy of information they contain is not always up to CA standards, their content is usually sufficient to permit the technical experts at CAS to prepare an adequate CA citation. But this form of patent handling is exceptional. All other patent abstracting is performed from the complete patent document.

Perhaps the primary problem facing the abstracting and indexing service, as well as the one in need of information, in dealing with worldwide patent literature is the fact that more than one document exists for each separate invention. It is common practice for inventors to patent their inventions in several countries. This practice, from the viewpoint of the abstracting and indexing service, is tantamount to duplicating a given article in a half dozen or more different journals. Mechanisms have been devised to identify for the user, through the CA Patent Index, these nests or families of patents covering the same invention, but it is a tremendous challenge to the abstracting and indexing service to identify this duplication in the first place, thereby avoiding the expenditure of time and intellectual effort that would otherwise be used to abstract and index the same thing many times over.

Thus far the diversity of language, transliteration, the need for bilingual comprehension, and the patent literature have

been presented as examples of foreign-literature challenges presented to CAS. Fundamental differences in scientific publishing procedures, and even in publication philosophy, among the various countries of the world provide additional challenges in the continuing search for completeness of coverage. Progress and interim research reports, extended abstracts, abstracts of papers read at meetings, preprints, and doctoral theses are just a few examples of document types, foreign and domestic, restricted or excluded by policy decision from CA coverage because of their incomplete and ephemeral nature and/or the fact that the same information invariably reappears in the recognized primary literature.

DOCUMENT ACCESSIBILITY

The CAS Source Index, more familiarly referred to by its acronym CASSI, was alluded to earlier. It is an important bibliographic information reference for the more than 50 000 scientific publications that comprise the world's scientific and technical literature as abstracted and indexed by CA over the past 75 years.

But perhaps CASSI's primary service is to the researcher whose scientific curiosity was piqued by a CA abstract and who now seeks the full document for details. The foreign-language literature has been a frequent obstacle to information gathering simply because of possible limited availability of the primary document. The purpose of the abstract is to alert and inform; it is not intended as a substitute for the original writing. Nevertheless, it is often made to serve as such, especially if the original is a none-too-widely distributed foreign publication. Through CASSI a total of 369 of the world's major resource libraries, 69 of which are located in 28 foreign countries, share their holdings information with the user community and make the scientific literature available through a variety of lending or photocopy arrangements.

Document accessibility is further enhanced by CAS's Document Delivery Service, which makes available most complete articles, patents, technical reports, or other documents cited in CAS publications since 1972 and to which CAS itself has access. Soviet document availability extends back to 1966. Thus, the abstract with its bibliographic and technical information retrieval power is all the more useful in serving as the bridge between the users and the generators of knowledge.

Communication remains as one intangible, yet impelling, force that feeds research and drives it forward for the ultimate betterment of mankind. Its effectiveness is challenged by language and geography, a challenge that *Chemical Abstracts* accepts anew each day as it enters upon the next 75 years of service to science and technology.

REFERENCES AND NOTES

- 1) Baker, D. B. "Recent Trends in Chemical Literature Growth." *Chem. Eng. News* **1981**, 59 (June 1), 29-34.
- 2) "Cyrillic Romanization"; American National Standards Institute: New York, 1976; ANSI Standard Z 39.24-1976.
- 3) "The Pinyin Chinese-English Dictionary"; Jingrong, Wu, Ed.; Beijing Foreign Language Institute: Beijing, China, 1979.
- 4) "Handbook for Pinyin Romanization of Chinese Proper Names", 1st ed.; Joint Publications Research Service: Arlington, VA, 1978.
- 5) "Japanese Romanization"; American National Standards Institute: New York, 1972; ANSI Standard Z 39.11-1972.
- 6) Rowlett, R. J., Jr. "Abstracts—Who Needs Them?". *NFAIS NewsL* **1981**, 23 (2), 26-36.
- 7) Pollick, P. J. "Patents and Chemical Abstracts Service". *Sci. Technol. Libr.* **1981**, 2 (2), 3-22.