Library and Information Science No. 14 1976

Romanization of Japanese Names in Chemical Literature

化学文献における日本人名のローマ字化

Masamitsu Negishi
根岸正光
Takeo Yamamoto
山本毅雄

要旨

科学論文の日本名をもつ著者が，その姓をいかにローマ字化しているかを統計的に研究し，姓によるローマ字化法の差について検討を加えた。オンライン情報検索システム TOOL-IR により，Chemical Abstracts Condensates 98 週分の著者名を検索した。ヘボン式と訓令式に差があり，どちらかを選択する場合，99％以上の著者がヘボン式を用いている。オ列長音のローマ字化には種々の方式があり，姓によりその選択は大きくことなるが，比較的多数が，ヘボン式，訓令式のどちらでもない“oh”を用いており，例えば‘太田’では80％以上が“Ohta”と綴っている。

1. Introduction

There are three major methods for the Romanization of Japanese language; the Hepburn method, the Kunrei method and the Nippon method. The last two are similar except for minor differences but there are some significant differences between the first and the other two methods. The choice among the methods has long been a matter of intense controversy.[1),2)] However, there have been few studies of the relative usage of the methods or of the other methods of Romanization in the actual situation.

In the present study, the methods of Romanization of last names of authors of scientific papers in chemistry are studied statistically. In chemistry, unlike some other fields of study, the practice of publishing papers in foreign languages is widespread among Japanese authors. The authors, even when they are publishing in Japanese, are conscious of their paper's being abstracted and of the possibility of the papers' circulation abroad. Under such circumstances, the authors are expected to Romanize their names in such a way as they wish the names to be known abroad.

An on-line bibliographic information system, TOOL-IR, was used for the study. In the following section, the system and the method of measurement will be described. The relative frequencies of usage of Hepburn vs. Kunrei methods will be compared where applicable, and other methods of Romanization will be identified. Some explanations for the observed phenomena will be presented.

Masamitsu Negishi, Faculty of Economics, The University of Tokyo.
根岸正光： 東京大学経済学部
Takeo Yamamoto, Computer Centre, The University of Tokyo.
山本毅雄： 東京大学大型計算機センター

2. Method of Measurement

The measurement was done using an on-line information retrieval system, TOOL-IR.[3),4)] The system has been in operation at the University of Tokyo Computer Centre, which is an interuniversity computer center for academic research since August 1974 and has been used by academic users from all over Japan since July 1975. Its database includes Chemical Abstracts Condensates (CACon), INSPEC Computer and Control Abstracts, and X-ray Crystallographic Data File from the Crystallographic Data Centre at the Cambridge University.[5)] Both batch and TSS searches are possible; there are more than 100 TSS terminals and 22 RJE stations connected to the center as of

Table 1. Comparison between Sakuma Ranking and the Present Result

Sakuma's Study		Present Study			Sakuma's Study		Present Study		
Name	Rank	Name Searched	Number of Reference	Rank	Name	Rank	Name Searched	Number of Reference	Rank
鈴木	1	Suzuki	4235	1	石川	28	Ishikawa, Isikawa	942	35
佐藤	2	Sato, Satou, Satoh, Satoo	3470	3	内田	29	Uchida, Utida	630	43
田中	3	Tanaka	3648	2	岡田	30	Okada	1173	24
山本	4	Yamamoto	3114	6	青木	31	Aoki	722	40
渡辺	5	Watanabe	2907	8	金子	32	Kaneko	622	44
高橋	6	Takahashi, Takahasi	3221	5	近藤	33	Kondo, Kondou, Kondoh, Kondoo	1255	22
小林	7	Kobayashi, Kobayasi	2969	7	阿部	34	Abe	1113	27
中村	8	Nakamura	2748	9	和田	35	Wada	1023	30
伊藤	9	Ito, Itou, Itoh, Itoo	3242	4	太田	36	Ota, Oota, Ohta, O'ta, Outa	1018	31
斎藤	10	Saito, Saitou, Saitoh, Saitoo	2165	11	小島	37	Kojima, Kozima, Ojima, Ozima	976	34
加藤	11	Kato, Katou, Katoh, Katoo	2537	10	島田	38	Shimada, Simada	544	46
山田	12	Yamada	2127	12	遠藤	39	Endo, Endou, Endoh, Endoo	694	42
吉田	13	Yoshida, Yosida	2049	14	田村	40	Tamura	913	36
佐々木	14	Sasaki	1300	21	高木	41	Takagi	755	38
井上	15	Inoue, I'nove, Iue	2058	13	中野	42	Nakano	736	39
木村	16	Kimura	1848	15	小山	43	Koyama, Oyama	560	45
松本	17	Matsumoto, Matumoto	1540	18	野田	44	Noda	450	50
清水	18	Shimizu, Simizu, Simidu, Shimidu	1563	17	福田	45	Fukuda, Hukuda	709	41
林	19	Hayashi, Hayasi	1847	16	大塚	46	Otsuka, Ohtsuka, Ootsuka, O'tsuka, Outsuka, Otuka, O'tuka, Ohtuka, Ootuka, Outuka	502	49
山口	20	Yamaguchi, Yamaguti	1412	20					
長谷川	21	Hasegawa	994	33					
小川	22	Ogawa, Kogawa	1124	26	岡本	47	Okamoto	1066	29
中島	23	Nakajima, Nakazima, Nakashima, Nakasima	1234	23	辻	48	Tsuji, Tuzi, Tsuzi, Tuji	533	48
山崎	24	Yamazaki, Yamasaki	1006	32	横山	49	Yokoyama	537	47
橋本	25	Hashimoto, Hasimoto	1079	28	後藤	50	Goto, Gotou, Gotoh, Gotoo	852	37
森	26	Mori	1458	19					
池田	27	Ikeda	1155	25					

November 1976.

Only the CACon database, corresponding to Chemical Abstracts Vol. 80, Issue 1 (first issue, 1974) to Vol. 84, issue 20 (April 17, 1976), was used for the present study. CACon magnetic tapes are converted to TOOL-IR database as they arrive. The following data elements are indexed as the searchable keywords: document title, keyword phrase, personal author name, CA section/subsection number, journal CODEN. The user may combine those keywords in a query to retrieve desired documents. In the present study, only the personal author names were searched.

Names were chosen based on the ranking of frequency of Japanese last names compiled by Sakuma.[6] First, the 50 most common names in the Sakuma ranking were used. As many ways of Romanization as possible were collected for each name. They were used as the search words in the TOOL-IR query, and the number of documents for each spelling of the last name was obtained. When necessary, other names were used to test the working hypotheses formulated.

There was no way of knowing whether an author of a retrieved paper is a Japanese or he is a foreign national of Japanese descent: no attempt was made to distinguish between the two. However, where there was a possibility of coincidence of a Japanese name with a last name of non-Japanese origin, the number of documents whose author has a middle name was not counted.

3. Results

3.1 Comparison with the Sakuma Ranking

A comparison of the present result with the Sakuma ranking is shown in Table 1. Sakuma ranking has been compiled by using the lists of elementary and junior high school teachers in all the prefectures in Japan as the primary sources of data, and the central government employees' list and several local phone books as auxiliary sources. In the ranking, variant characters for the same sound are distinguished

Table 2. Usage of the Hepburn and the Kunrei Methods

Name (Hepburn)	Number of References (Percentage)		
	Hepburn	Kunrei	Total
Takahashi	3207(99.6)	14(0.4)	3221(100)
Kobayashi	2953(99.5)	16(0.5)	2969(100)
Yoshida	2044(99.8)	5(0.2)	2049(100)
Shimizu	1555(99.5)	8(0.5)	1563(100)
Hayashi	1844(99.8)	3(0.2)	1847(100)
Nakajima	214(99.1)	2(0.9)	216(100)
Hashimoto	1071(99.4)	7(0.6)	1079(100)
Ishikawa	942(100.0)	0	942(100)
Shimada	544(100.0)	0	544(100)
Sub Total ("shi")	14375(99.6)	55(0.4)	14430(100)
Nakajima	1013(99.5)	5(0.5)	1018(100)
Kojima	876(97.7)	21(2.3)	897(100)
Ojima	74(93.7)	5(6.3)	79(100)
Tsuji	528(99.1)	5(0.9)	533(100)
Sub Total ("ji")	2491(98.6)	36(1.4)	2527(100)
Matsumoto	1523(98.9)	17(1.1)	1540(100)
Ōtsuka	502(100.0)	0	502(100)
Tsuji	527(98.9)	6(1.1)	533(100)
Sub Total ("tsu")	2552(99.1)	23(0.9)	2573(100)
Yamaguchi	1410(99.9)	2(0.1)	1412(100)
Uchida	626(99.4)	4(0.6)	630(100)
Sub Total ("chi")	2036(99.7)	6(0.3)	2042(100)
Fukuda	705(99.4)	4(0.6)	709(100)
Grand Total	22159(99.4)	124(0.6)	22283(100)

Table 3. Usage of the Hepburn and the Kunrei Methods (Names with "fu")

Name (Hepburn)	Number of References (Percentage)		
	Hepburn	Kunrei	Total
Fukuda	132(100.0)	0(0.0)	132(100)
Fujii	185(99.5)	1(0.5)	186(100)
Fukushima	90(100.0)	0(0.0)	90(100)
Fujiwara	104(100.0)	0(0.0)	104(100)
Total	511(99.8)	1(0.2)	512(100)

whereas variant readings for the same character are not. As in the present study only Romanized name spellings are distinguished and only the number of papers with the particular author as one of the coauthors is obtained (but not the number of persons with the last name), there are some inevitable discrepancies. However, the general agreement between the two methods seems to be good. The rank correlation coefficient for the first 20 entries of the table was calculated to be 0.93.

The above result shows that, despite the difference between the methods and between the scope of the two studies both of them represent a fair sampling of frequencies of Japanese last names. Sakuma has also reported an estimated population for each Japanese last name. It is, however, in poor agreement with the present result: it seems that Sakuma's population distribution is too low for more common names and too high for less common names.

3.2 Ratios of Usage of the Hepburn and the Kunrei Methods

Tables 2 and 3 show frequencies of usage of different methods of Romanization for last names shown in Table 1. The differences are caused by last names containing sounds "ō", "shi", "ji", "tsu", "chi" and "fu" (hereafter the Hepburn method will be used in the present report to designate the Japanese sound). As there are various methods used for Romanizing "ō", they will be discussed separately in section 3.3. Here, the ratios of usage of the Hepburn and the Kunrei methods for the other five sounds will be discussed.

Table 2 shows the ratios of the usage of the two methods (among the names listed, the Nippon method is identical with the Kunrei method) for the top 50 names listed in Table 1. From Table 2 it is clearly seen that Hepburn method is chosen by the majority of the present Japanese chemists. Among the Kunrei spellings, "zi" (1.6%) seems to be slightly more popular than "tu" (0.9%), "hu" (0.6%), "si" (0.4%) and "ti" (0.3%). However, the size of the sample is not large enough for a more detailed discussion.

As Fukuda/Hukuda was the only example of a last name with "fu" in the top 50 names, other names with the sound, "Fujii" (52nd), "Fukushima" (92nd) and "Fujiwara" (99th)

Table 4. Various Spellings of "ō" (Top 50 Names)

Name (Hepburn)	Number of Reference (Percentages)				
	"o"	"oh"	"ou"	"oo"	Total
Satō	3260 (93.9)	203 (5.9)	7 (0.2)	0 (0.0)	3470 (100)
Katō	2442 (96.3)	91 (3.6)	3 (0.1)	1 (0.0)	2537 (100)
Gotō	791 (92.8)	58 (6.8)	3 (0.4)	0 (0.0)	857 (100)
Sub Total	6493 (94.7)	352 (5.1)	13 (0.2)	1 (0.0)	6859 (100)
Kondō	1247 (99.4)	8 (0.6)	0 (0.0)	0 (0.0)	1255 (100)
Endō	652 (93.9)	35 (5.0)	4 (0.6)	3 (0.4)	694 (100)
Itō	2783 (85.8)	446 (13.8)	12 (0.4)	1 (0.0)	3242 (100)
Saitō	2080 (96.1)	76 (3.5)	8 (0.4)	1 (0.0)	2165 (100)
Total ("xō" type)	13255 (93.2)	917 (6.5)	37 (0.3)	6 (0.0)	14215 (100)
Ōta	173 (17.0)	788 (77.4)	1 (0.1)	56 (5.5)	1018 (100)
Ōtsuka	261 (52.0)	227 (45.2)	0 (0.0)	14 (2.8)	502 (100)
Total ("ōx") type	434 (28.6)	1015 (66.8)	1 (0.1)	70 (4.6)	1520 (100)
Grand Total	13689 (87.0)	1932 (12.3)	38 (0.2)	76 (0.5)	15735 (100)

were searched from CA Condensates Vol. 84. The result, shown in Table 3, also confirms the conclusion that most Japanese chemists spell their names in the Hepburn method when he must choose between the Kunrei and Hepburn methods.

3.3 Various Spellings of "ō"

Table 4 shows the frequencies of various spellings for the sound "ō", in the top 50 names in the Sakuma ranking. It is seen from the table that several alternatives are used in place of the macron (¯) which is the standard symbol for designating a long vowel in both the Hepburn and the Kunrei methods. As the macron is often unavailable in publications and abstract services outside Japan, the authors are left to devise the alternative spelling by himself.

In Table 4, an obvious difference lies between names with "ō" at the first syllable (hereafter to be called as "ōx" type names) and those with "ō" at the end (hereafter called "xō" type names). In all the "xō" type names, the spelling "o" is most frequently used (average: 93.2%), followed by "oh" (6.5%) and "ou" (0.3%). There were negligible instances of "oo". In "ōx" type names, on the other hand, "oh" (66.8%) is most commonly used, followed by "o" (26.6%), "oo" (4.6%) and finally "ou" (0.1%). Among the "xo" type

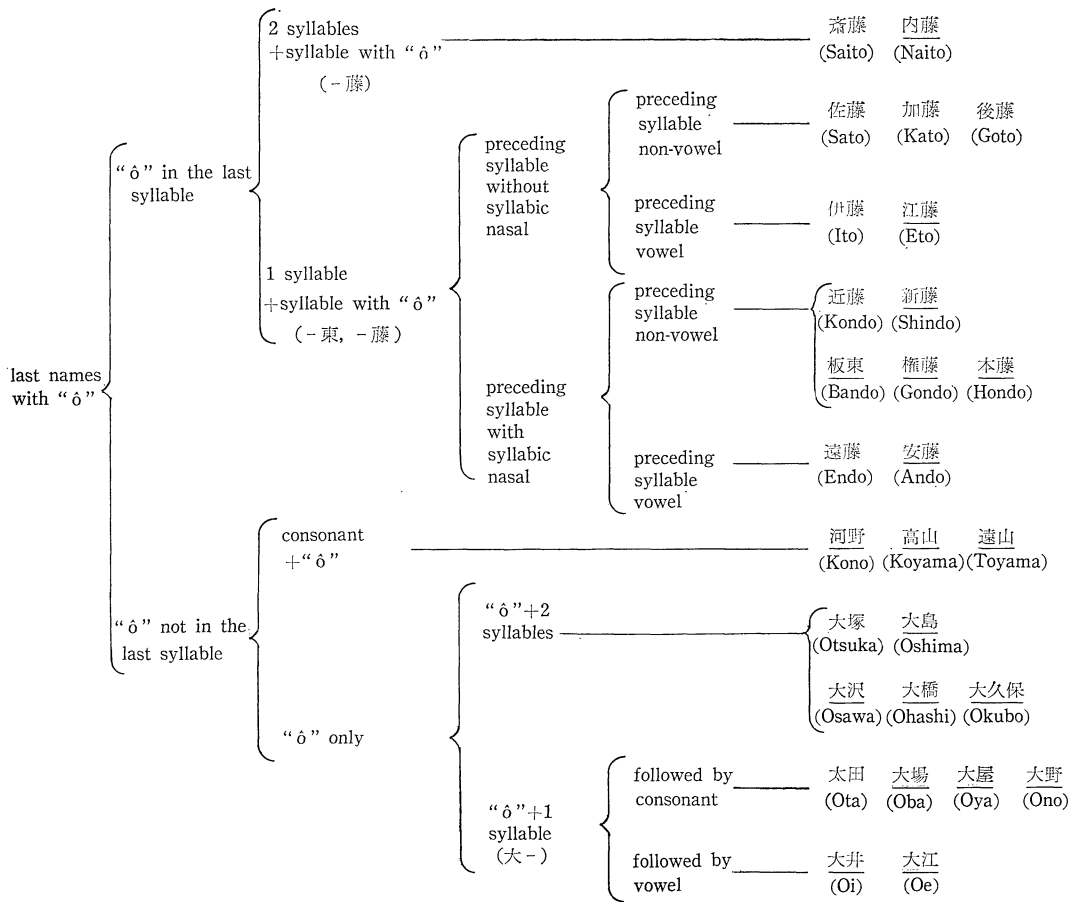

Fig. 1. Phonetic Classification of Japanese Last Names with Sound "ô"

Table 5. Various Spellings of "ō" ("xō" type)

Name (Hepburn)	Number of References (Percentage)				
	"o"	"oh"	"ou"	"oo"	Total
Kondō	206 (100.0)	0	0	0	206 (100)
Shindō	34 (100.0)	0	0	0	34 (100)
Gondō	11 (84.6)	2 (15.4)	0	0	13 (100)
Bandō	10 (100.0)	0	0	0	10 (100)
Hondō	2 (100.0)	0	0	0	2 (100)
Sub Total (Kondō)	263 (99.2)	2 (0.8)	0	0	265 (100)
Endō	127 (96.2)	5 (3.8)	0	0	132 (100)
Andō	100 (98.2)	2 (2.0)	0	0	102 (100)
Sub Total (Endō Type)	227 (97.0)	7 (3.0)	0	0	234 (100)
Itō	430 (85.5)	70 (13.9)	3 (0.6)	0	503 (100)
Etō	21 (100.0)	0	0	0	21 (100)
Sub Total (Itō type)	451 (86,1)	70 (13.4)	3 (0.6)	0	524 (100)
Saitō	392 (97.8)	9 (2.2)	0	0	401 (100)
Naitō	71 (97.3)	2 (2.7)	0	0	73 (100)
Sub Total (Saitō type)	463 (97.8)	11 (2.3)	0	0	474 (100)

Table 6. Various Spellings of "ō" ("ōx" Type)

Name (Hepburn)	Number of References (Percentage)				
	"o"	"oh"	"ou"	"oo"	Total
Ōta	23 (11.7)	164 (83.2)	1 (0.5)	9 (4.6)	197 (100)
Ōba	14 (34.1)	22 (53.7)	0 (0.0)	5 (12.2)	41 (100)
Ōya	10 (52.6)	9 (47.4)	0 (0.0)	0 (0.0)	19 (100)
Ōno (Ono)	151 (57.9)	105 (40.2)	0 (0,0)	5 (1.9)	261 (100)
Ōi	5 (20.0)	17 (68.0)	0 (0.0)	3 (12.0)	25 (100)
Ōe	12 (57.1)	8 (38.1)	0 (0.0)	1 (4.8)	21 (100)
Ōtsuka	41 (43.6)	52 (55·3)	0 (0.0)	1 (1.1)	94 (100)
Ōshima	22 (24.7)	61 (68.5)	0 (0.0)	6 (6.7)	89 (100)
Ōsawa	25 (37.9)	34 (51.5)	0 (0.0)	7 (10.6)	66 (100)
Ōhashi	24 (61.5)	13 (33.3)	0 (0.0)	2 (5.1)	39 (100)
Ōkubo	25 (33.8)	39 (52.7)	0 (0.0)	10 (13.5)	74 (100)
Sub Total (Ōtsuka)	137 (37.8)	199 (55.0)	0 (0.0)	26 (7.2)	362 (100)
Kōno	46 (80.7)	11 (19.3)	0 (0.0)	0 (0.0)	57 (100)
Kōyama (Koyama)	88 (95.7)	4 (4.3)	0 (0.0)	0 (0.0)	92 (100)
Tōyama (Toyama)	36 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	36 (100)

names, it is noticed that the occurrence of "oh" in "itoh" is exceptionally high (13.8%) compared to the occurrence of the spelling in other names.

To study the differences in more detail, more names with "ō" were collected and classified by their phonetic characteristics as in Fig. 1. Here, underlined names indicate names outside the top 50 of the Sakuma ranking. The frequencies of these names are searched in CA Condensates Vol. 84 as before and are listed in Table 5 and 6. Though the numbers are too small to draw a firm conclusion, some additional observations may be made from these tables: Among "xō" type names,

a. "Itō" has an unusually high frequency of "oh".

b. "Kondō" has a low frequency of "oh".

Among "ōx" names,

a. "Kōno" type has distinctively low frequencies of "oh" compared to other types.

b. "Ōhashi" has a relatively low frequency of "oh" (33.3%) compared to the other names of "Ōtsuka" type.

3.4 Other Spellings

Other examples of last names with variant spellings in the current data base were found to be too small to draw a reliable conclusion. Some preliminary searches are made:

"jō"....."Chujō" (5 cases) and "Hojō" (12 cases) were all spelled as CHUJO and HOJO. Spellings such as CYUZYO or HOZYO were absent.

"n"....."Nanbu" was spelled as NAMBU in 8 cases and as NANBU in 6 cases. There was one JIMBO and one JINBO for "Jinbo".

4. Discussion

The popularity of the Hepburn method and the widespread use of "oh" for "ō" indicate that most of Japanese chemists have chosen spellings of their Romanized names so that, pronounced like English, the original sound of the name may be approximated. The choice seems to have been made pragmatically, with little care for the regulations or established rules of Romanization if they are impractical. Most authors of name Ōta are presumably aware of the fact that, if they spelled their names according to the rule, the macron will be taken off once their articles are abstracted, reviewed or cited by other authors.

Differences in the usage of various spellings for the sound "ō" among several types of names are perhaps a material for a psycological study. Here, a few possible explanations will be presented.

a. The popularity of "oh" among names beginning with the character "大" or "太", such as "Ōta", "Ōtsuka" and "Ōno", may be explained by the similarity of the spelling with the historical phonetic expression (rekishiteki kanazukai) for those characters, "おほ (oho)-". In the case of "Ōhashi", "Ohhashi", is understandably less popular.

b. Likewise, the (current) phonetic expression "-とう (-tou)" and "-どう (-dou)" for names ending with "藤" may explain the relative popularity of "ou" in "xō" type names compared to that among "ōx" type names.

c. Other factors such as the number of syllables and the numbers of alphabetic characters in the name, the position of the syllable in question, the position of the accent and the nature of the adjacent syllables seem to influence the preference among various spellings. More data is needed to prove these points, however.

d. Authors are likely to be influenced by available precedence in choosing the spelling. The fact that "Itō" is much more often spelled with "oh" than similar names may, for example, have been influenced by the well-known name of a company, "C. ITOH".

Acknowledgment

The authors' thanks are due to Prof. Shizuo Fujiwara and Prof. Takehiko Shimanouchi of the University of Tokyo for their support for the study. They also appreciate the technical

assistance of Mr. Mamoru Ushimaru in developing TOOL-IR. The study was supported by the Ministry of Education, Science and Culture fund Tokutei Kenkyu (I), No. 111730.

References

1) Cabinet Notification No. 1, "Rômazi no Tuzurikata (Spelling of Rômazi)," Dec. 9, 1954, p. 321–328.
2) Ootuka, H. and E. Naito, "Scope of ISO/TC 46–Documentation, the 16th Plenary Meeting (in Japanese)," *Dokumentêsyon Kenkyu*, vol. 26, no. 8/9, 1976, p. 321–328.
3) Yamamoto, T., M. Negishi, M. Ushimaru, Y. Tozawa, K. Okabe and S. Fujiwara, "TOOL-IR: An On-Line Information Retrieval System at an Inter-University Computer Center." *Proc. 2nd US-Japan Computer Conf.*, 1975, p. 159–165.
4) Negishi, M. and T. Yamamoto, "Man-Machine Interface in TOOL-IR, an On-Line Bibliographic Information Retrieval System," *Information Processing in Japan.* (in press)
5) Shimanouchi, T. and T. Yamamoto, "Crystallographic Data Services in Japan," *Proc. 5th Biennial CODATA Conference.* (in press)
6) Sakuma, E., *Nipponjin no Sei* (*Last Names of Japanese*). Tōkyo, Rikugei Shobō, 1972.