DCU Ollscoil Chathair
Bhaile Átha Cliath
Dublin City University

ローマ字表記とその先のローマ字

Grapholinguistics in the 21st Century 2024
Università Ca' Foscari, Venice, Italy, 23-25 October 2024

Rōmaji for romanisation and beyond:
A non-phonemic adaptation of the Latin script in Japanese

Keisuke Honda | 本田 啓輔

keisuke.honda@dcu.ie

Dublin City University | Ollscoil Chathair Bhaile Átha Cliath
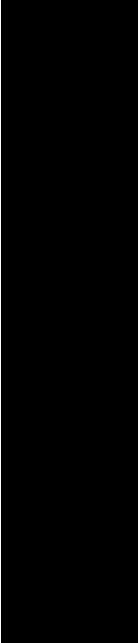
About this study

Rōmaji

❖ Latin script adapted in the Japanese writing system (JWS)

❖ Limited but increasing use + ongoing orth. revision

❖ Commonly described as phonemic / segmental / alphabetic
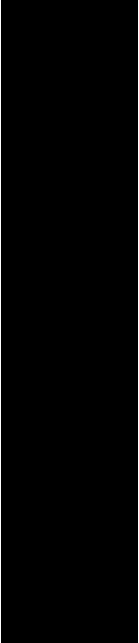
❖ Definition and scope issues (Honda 2023)

Questions

❖ Is rōmaji phonemic?

❖ Is the Latin script intrinsically phonemic?

Disclaimer: Focus on grain size, not phonemicity or orth. depth.

1. Latin script and phonemic writing

2. Rōmaji as a component script

3. Romanisation and beyond

4. Conclusion

Latin script

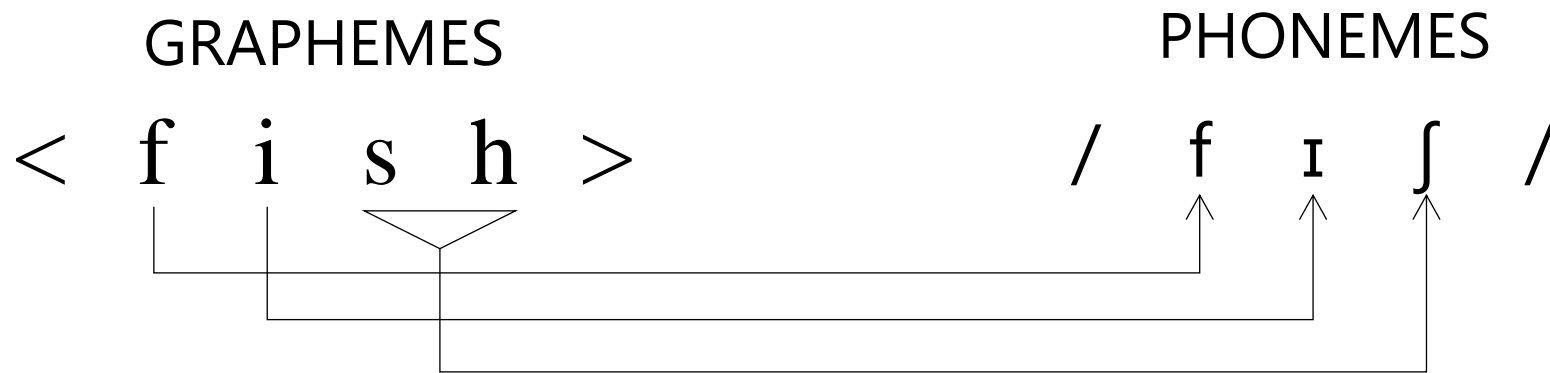## Brief outline

❖ Originally developed for Latin

❖ Later adapted for many languages of the world

❖ In principle, 1 grapheme → 1 phoneme

E.g. English <fish>

GRAPHEMES   PHONEMES

< f i s h >   / f ɪ ʃ /

Grapheme: Minimal graphematic unit (monographic / polygraphic)
Phoneme: Minimal phonological unit (segmental / suprasegmental)
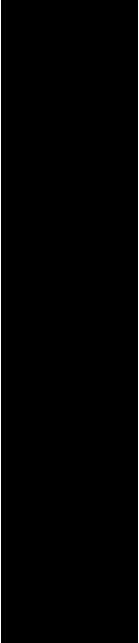
Linguistic characterisation

Common observation

❖ Distinction of primary and secondary mapping types (e.g. Gelb 1963)

❖ Focus on phonemic mapping (e.g. Daniels & Bright 1996)

❖ Equation with phonemic writing (e.g. Unicode 2024)

E.g. Latin script in the English writing system

❖ Phonemic:
                    <fish> → /fɪʃ/
                    <cat> → /kæt/
                    <AIDS> → /eɪdz/

'Primary' type in most words

❖ Non-phonemic:
                    <HIV> → /eɪtʃ aɪ viː/
                    <III> → /θɹiː/
                    <lb> → /paʊnd/

'Secondary' type in some words

Japanese writing system

Brief outline

❖ Mixed use of multiple scripts

❖ Assignment of specific functions to each script

❖ Limited but increasing use of romaji (Irwin 2011; Tomoda 2005; Yamashita 2019)

E.g. <ＤＶＤでＡｋｉｒａのアニメ映画を見た。>

RŌMAJI	HIRAGANA	RŌMAJI	HIRAGANA	KATAKANA	KANJI	HIRAGANA	KANJI	HIRAGANA	PUNCT.
ＤＶＤ	で	ＡＫＩＲＡ	の	アニメ	映画	を	見	た	。
dībuidī	*de*	***Akira***	*no*	*anime*	*eiga*	*o*	*mi*	*ta*	
DVD	=INSTR	**Akira**	GEN	animation	film	=ACC	see	-PAST.AFF.PLAIN	

'Watched the animation film *Akira* on **DVD**.'

Rōmaji

Common assumption

❖ Mainly used for romanisation (e.g. Smith 1996; Lunde 2009; Joyce & Masuda 2018)

❖ Romanisation: Conversion of a non-Latin-based writing system into the Latin script through transcription or transliteration (Coulmas 1996)

E.g. Rōmaji in actual use
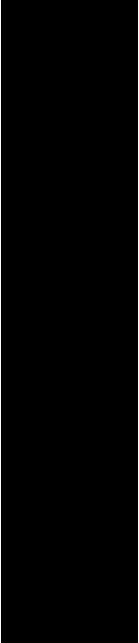
Source: Wikimedia Commons

Source: brax.com

Source: X @SN_20200122

Rōmaji use types

Classification (based on Irwin 2011, Haralambous 2023, et al.)

	Loanword	SRE word*	JP word	Phonemic
1. Romanisation	—	—	Tokyo	✓
2. Acronym	SIM	SMAP	KAT-TUN	(✓)
3. Graphic foreignism	Google	Radwimps	Mazda	(✓)
4. Graphic abbreviation	TV	gj	w	✗
5. Initialism	VR	CM	NHK	✗
6. Letter morpheme	X (線)	Y (シャツ)	VIO	✗
7. Xenography	AM	V	—	✗
8. Hybrid	XR	DX	HIJK	(✓) / ✗

* SRE: Semantically Remodelled English = English-based coinage in Japanese (Irwin 2011)

Phonemic uses

Romanisation

Conversion of a non-Latin script into the Latin script through transcription or transliteration.

Examples

<Tokyo>	<Akira>	<Fujitsu>
/tōkyō/	/akira/	/fujitsū/
ʻPLACE NAMEʼ	ʻPERSON NAMEʼ	ʻCOMPANY NAMEʼ
Jpn. 東京 / とうきょう	Jpn. 明, 昭, 昌, etc.	Jpn. 富士通

→ **Phonemic (segmental, contrastive)**

Acronym
(after Haralambous 2023)

Initial letters of compound elements pronounced as a single word.

Graphic foreignism
(after 'foreignism')

Loanword in the original spelling or JP word in unconventional romanisation.

Examples

<SIM>
/shimu/
'SIM'

Eng. *subscriber identity module*

<KAT-TUN>
/katūn/
'BOYBAND NAME'

Jpn. name initials

Examples

<Google>
/gūguru/
'COMPANY NAME'

Eng. *Google*

<Mazda>
/matsuda/
'COMPANY NAME'

Jpn. マツダ
+ Ave. *Ahura Mazdā*

→ **Phonemic + Polyphonemic (epenthesis / idiosyncratic spelling)**

Graphic abbreviation
(after Haralambous 2023)

Graphically shortened word pronounced in the original long form.

Initialism
(after Haralambous 2023)

Initial letters of compound elements pronounced as letter name sequence.

Examples

<TV>
/terebi/
'TV'

Eng. *television*

<w>
/warai/
'laughter, LOL'

Jpn. 笑い

Examples

<VR>
/bui āru/
'VR'

Eng. *virtual reality*

<NHK>
/enu etchi kē/
'COMPANY NAME'

Jpn. 日本放送協会
Nippon Hōsō Kyōkai

→ **Morphographic (phonographically derived)**

Letter morpheme
(after Riha & Baker 2010)

A letter name behaving as a unique morpheme.

Examples

<X(線)>
/ekkusu(sen)/
'X(-ray)'

Eng. *X-ray*

<VIO>
/bui ai ō/
'pubic hair'

Letter shape
iconicism

Xenography
(after Coulmas 1996)

Borrowed written word pronounced in JP translation.

Examples

<A.M.>
/gozen/
'A.M.'

<V>
/yūshō/
'victory'

Eng. *A.M.*
< Lat. *ante meridiem*

Eng. *victory*

→ **Morphographic (ideo-/picto-/phonographically derived)**

Partially phonemic / non-phonemic uses

Hybrid

Combination of different rōmaji use types.

Examples

<XR>
/ekkusu āru/
'XR'

<DX>
/dī ekkusu/
'digital transformation'

<HIJK>
/etchi ai jē kē/
'marriage after baby'

Eng. *extended reality*;
<X> for *ex-*

Eng. **d**igital + **trans**formation;
<X> for *trans-* 'across'

Jpn. H *etchi* 'sex'
+ 愛 *ai* 'love' + Eng. *junior*
+ Jpn. 結婚 *kekkon* 'marriage'

→ **Phonemic + Polyphonemic + Morphographic**

Rōmaji use types

Classification (based on Irwin 2011, Haralambous 2023, et al.)

	Loanword	SRE word*	JP word	Phonemic
1. Romanisation	—	—	Tokyo	✓
2. Acronym	SIM	SMAP	KAT-TUN	(✓)
3. Graphic foreignism	Google	Radwimps	Mazda	(✓)
4. Graphic abbreviation	TV	gj	w	✗
5. Initialism	VR	CM	NHK	✗
6. Letter morpheme	X (線)	Y (シャツ)	VIO	✗
7. Xenography	AM	V	—	✗
8. Hybrid	XR	DX	HIJK	(✓) / ✗

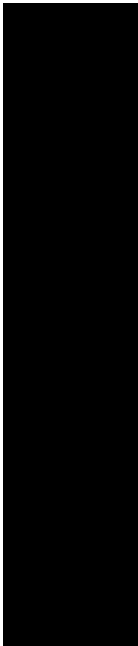* SRE: Semantically Remodelled English = English-based coinage in Japanese (Irwin 2011)

Distribution of rōmaji use types

Corpus-based survey (questioned by Terry Joyce, p.c.)

❖ Material: Core subset of the 100-million-word Balanced Corpus of Contemporary Written Japanese (BCCWJ: Maekawa et al. 2014)

❖ Method: 1) Sampled 10% (n = 798) of all words containing 1+ rōmaji
2) Sorted items using BCCWJ's parameters (e.g. Word Type)
3) Checked the use type of each item (e.g. romanisation)

❖ Result: 16.04% of tokens identified as romanisation (i.e. phonemic)

	D	E	F	L	O	Q
1	前文脈	キー	後文脈	品詞	書字形	語種
2	と|なっ|て|いる#スタッフ|の|想う|理想|の|	RV	|ライフ#鵜飼|英一#　|キャンバー|を|買っ|て	名詞-普通名詞-一般	R V	記号
3	|の|時計|って|珍しい|か|な、|と|思っ|て|	VIVIER	|で|購入|。#¥|一|万|二千|と|安かっ|た|！|	名詞-固有名詞-一般	V I V I E R	固
4	644655#ピンク|の|T|シャツ|w|w|w|	w	|w|w|w|w|w|w|w|w|w|w|w|w|w|w|w|w	記号-文字	w	記号
5	と|で|意見|一致#・|国際|熱|核|融合|炉|（|	ITER	|）|が|成功|裡|に|協力|が|行わ|れ|て|いる|	名詞-普通名詞-一般	I T E R	記号
6	|対決|を|行う|。#特集|歌謡|コンサート#（|	NHK	|＝|後|8|・|O|）#　|これ|まで|に|放送|さ	名詞-固有名詞-一般	N H K	固
7	|に|入る|メール|が|全部|ウェブ|メール|の|	Yahoo		|メール|に|入っ|て|いる|こと|を|一昨|日	名詞-固有名詞-一般	Y a h o o	固
8	弐|で|、|DF|金|泰映|を|下げ|、|ベテラン|	FW	|の|黄|善洪|を|投入|し|た|の|だ|。#　|スベ	名詞-普通名詞-一般	F W	記号
9	|IN|　|THE|　|SHOW|　|二千|三#	THE		|PRAYER|—|私|だけ|の|輪舞|が|誕	名詞-普通名詞-一般	T H E	外
10	後|7|・|二十|）#　|ドイツ|の|ハノーバー|	NDR	|フィルハーモニー|が|、|6|月|に|首席|指揮	名詞-固有名詞-一般	N D R	固

This study

Questions

❖ Is rōmaji really phonemic?

❖ Is the Latin script intrinsically phonemic?

→ **NO** (or, is it?)

Future research

❖ More elaborate analysis using another dataset / method

❖ Latin script in other non-Latin-based writing systems

(e.g. *luómǎ zìmǔ* in Chinese – *pīnyīn* and other uses)

References

Coulmas, Florian. 1996. *The Blackwell encyclopedia of writing systems*. Oxford & Malden: Blackwell Publishers.

Fukumori, Takahiro [福盛貴弘]. 2003. *Nihongo no moji no bunruian* [日本語の文字の分類案] / *A classification of Japanese graphs*. *Ippan Gengogaku Ronsō* [一般言語学論叢] / *Journal of General Linguistics* 6. 1-16.

Gelb, Ignatius J. 1963. *A study of writing*. Chicago & London: University of Chicago Press.

Haralambous, Yannis. 2020. Grapholinguistics, TEX, and a June 2020 conference. *TUGboat* 41. 12-19. http://www.tug.org/TUGboat/tb41-1/tb127haralambous-grapholinguistics.pdf. (Last accessed on 14th October 2024)

Haralambous, Yannis. 2023. Abbreviations from a grapholinguistic point of view. A paper in printing.

Hattori, Shirō [服部四郎]. 1979. On'inron to seishohō: Shin Nipponshiki tsuzurikata no teishō [音韻論と正書法 - 新日本式つづり方の提唱 'Phonology and orthography: A proposal for a new Japanese style spelling]. Tokyo: Taishūkan Shoten [大修館書店].

Honda, Keisuke. 2023. How alphabetic is Japanese rōmaji? Implications for the description and classification of writing systems. A conference paper presented at *the Association for Written Language & Literacy's 14th International Workshop on Writing Systems and Literacy* (Temple University Rome, Italy, 10-12 November 2023).

Irwin, Mark. 2011. *Loanwords in Japanese.* Amsterdam & Philadelphia: John Benjamins Publishing Company.

Joyce, Terry. & Masuda, Hisashi. 2018. Introduction to the multi-script Japanese writing system and word processing. In Pae, Hye. K. (ed.), *Writing systems, reading processes, and cross-linguistic influences: Reflections from the Chinese, Japanese and Korean languages*, pp. 179-200. Amsterdam: John Benjamins Publishing Company.

Kojima, Eiko [小嶋栄子]. 2001. Nihon ni okeru rōmaji no rekishi [日本におけるローマ字の歴史] (The history of rōmaji in Japan). In Nihongo Bumpō Kenkyūkai Kenkyūkai Hōkoku [日本語文法研究会 研究会報告] (Research Group Report of the Japanese Grammar Research Group) 22. 19-40.

López Rúa, Paula. 2004. Acronyms & Co.: A typology of typologies. *Estudios Ingleses de la Universidad Complutense* 12, pp. 109-129.

Lunde, Ken. 2008. *CJKV information processing*. Sebastopol: O'Reilly Media, Inc.

Maekawa, Kikuo et al. 2014. Balanced Corpus of Contemporary Written Japanese. *Language Resources and Evaluation*, 48(2), pp. 345-371.

Meletis, Dimitrios. 2018. *The nature of writing. A theory of grapholinguistics*. Brest: Fluxus Editions.

References

MoE = Ministry of Education in Japan [文部省]. 1950. *Rōmaji kyōiku no shishin* [ローマ字教育の指針] / *A guide to the romaji education*. Tokyo: Ministry of Education [文部省].

Neef, Martin. 2015. Writing systems as modular objects: Proposals for theory design in grapholinguistics. *Open Linguistics* 1. 708-721.

Riha, Helena & Baker, Kirk. 2010. Letter words: Using Roman letters to create words in Chinese. In Rainer, Franz, Dressler, Ddieter Dieter Kastovsky & Luschutzky, Hans Christian (eds.), *Variation and change in morphology: Selected papers from the 13th International Morphology Meeting, Vienna, February 2008*, pp. 193-200. Amsterdam & Philadelphia: John Benjamins Publishing Company.

Sasahara, Hiroyuki [笹原宏之]. 2001. Nihon no moji [日本の文字] / Japanese writing system. In Kōno, Rokurō [河野六郎]., Chino, Eiichi [千野栄一] & Nishida, Tatsuo [西田龍雄] (eds.), *The Sanseidō encyclopaedia of linguistics vol. 7: Scripts and writing systems of the world*, pp. 696-706. Tokyo: Sanseido Press [三省堂書店].

Satake, Hideo [佐竹秀雄]. 2005. Gendai Nihongo no moji to shokihō [現代日本語の文字と書記法] ('Modern Japanese writing and notation). In Hayashi, Chikafumi [林史典] (ed.), *Moji, shoki* [文字・書記] ('Writing, notation'), pp. 22-50. Tokyo: Asakura Shoten [朝倉書店].

Smith, Janet Shibamoto. 1996. Japanese writing. In Daniels, Peter T. & Bright, William. (eds.), *The world's writing systems*, pp. 209-217. New York: Oxford University Press.

Taylor, Insup. & Taylor, M. Martin. 2014. *Writing and literacy in Chinese, Korean and Japanese* (rev. ed.). Amsterdam & Philadelphia: John Benjamins Publishing Company.

Tomoda, Takako. 2005. Change in script usage in Japanese: A longitudinal study of Japanese Government White Papers on Labor. *Electronic Journal of Contemporary Japanese Studies* 5(2). https://www.japanesestudies.org.uk/articles/2005/Tomoda.html (last accessed on 7th November 2023)

Unicode Consortium. 2024. *Unicode® 16.0.0*. https://www.unicode.org/versions/Unicode16.0.0/ (last accessed on 22nd October 2024)

Yamashita, Yōko [山下洋子]. 2019. *Arufabetto ni yoru ryakki* [アルファベットによる略記] ('Abbreviations in the alphabet'). Tokyo: NHK Broadcasting Research Institute. https://www.nhk.or.jp/bunken/research/kotoba/20190601_5.html (Last accessed on 14th October 2024)

< 3Q >

/san kyū/

'thank you'

(Source:)

Appendix: A brief history of rōmaji

New addition to JWS

- Introduced in 16th century, increasingly common after mid-20th century
- Used for JP romanisation and foreign language education
- Different romanisation conventions developed over time
- Unified guidelines being discussed (Takeda 2023)

First introduction (1549-1616)

Brought in by Christian missionaries, used for language documentation, education and printing.

Second introduction (1854-1941)

EN and other Western languages taught in school as part of modern-isation process.

| 1500 CE | 1600 | 1700 | 1800 | 1900 | 2000 | |

First disruption (1639-1854)

Limited to trade and scholarly uses after the shogunate banned contact with the outside world.

Increased use (1945-Present)

Used by wider population for education, romanisation, inputting and other purposes.

Beginning of JWS (4-10th century CE)

Kanji introduced, kana invented.

Second disruption (1941-1945)

Banned during the Pacific War.