

Leonardo F. S. Boiko

## **O sistema de escrita japonês: Além da fala**

**São Paulo**

**6 de janeiro de 2017**



Leonardo F. S. Boiko

## **O sistema de escrita japonês: Além da fala**

Dissertação de mestrado (versão corrigida)

Universidade de São Paulo

Faculdade de Filosofia, Letras e Ciências Humanas

Programa de Pós-Graduação em Língua, Literatura e Cultura Japonesa

Orientadora: Leiko Matsubara Morales

São Paulo

6 de janeiro de 2017

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catalogação na Publicação  
Serviço de Biblioteca e Documentação  
Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo

B678s Boiko, Leonardo Ferreira da Silva  
O sistema de escrita japonês: Além da fala /  
Leonardo Ferreira da Silva Boiko ; orientadora Leiko  
Matsubara Morales. - São Paulo, 2016.  
163 f.

Dissertação (Mestrado)- Faculdade de Filosofia,  
Letras e Ciências Humanas da Universidade de São  
Paulo. Departamento de Letras Orientais. Área de  
concentração: Língua, Literatura e Cultura Japonesa.

1. Escrita (linguística). 2. Língua japonesa. 3.  
Língua chinesa. 4. Estatística. 5. Linguística. I.  
Morales, Leiko Matsubara, orient. II. Título.

*Dedico este trabalho à comunidade de leitores e escritores japoneses, que, ao optar por manter seu sistema de escrita vivo, enriquecem a cultura mundial com formas únicas de expressão e criatividade.*



# Agradecimentos

À minha orientadora, Leiko Matsubara Morales, pela confiança e apoio nas horas difíceis, e a meus colegas orientandos;

Aos profs. John Defrancis, Marshall Unger, Victor Mair e outros que, embora tenham chegado a conclusões opostas à minha sobre o valor da escrita *kanji*, produziram trabalhos de altíssimo nível, e que foram a minha formação no assunto; e em especial aos profs. Victor Mair e Marshall Unger, pela atenção, cortesia e gentileza nos contatos pela Internet;

A todos os professores que tive na USP, fontes de entusiasmo tanto quanto de conhecimento, e em particular a Paulo Chagas e Sylvio Horta, que muito me ensinaram sobre sistemas de escrita, despertando o desejo de aprender cada vez mais; e ainda ao prof. Chagas pela revisão detalhada desta dissertação, apontando correções em número tal que sua listagem tornou-se inviável;

Ao prof. Jim Breen, cujos dicionários colaborativos gratuitos são um recurso essencial para todos os aprendizes e pesquisadores de japonês, e que graciosamente orientou a escolha de *software* usado neste trabalho;

A todos os comentadores do meu *weblog*,<sup>1</sup> e em particular a Matt Treyvauld, Bathrobe, Flow e Rurōsha, pelas informações, discussões e apoio;

Aos japoneses que, ao saberem do tema deste trabalho, invariavelmente apoiaram-no calorosamente;

À comunidade de *software* livre, sem cuja dedicação e generosidade este trabalho—produzido em Debian GNU/Linux, X<sub>Y</sub>L<sup>A</sup>T<sub>E</sub>X, Python, R, Junicode, Emacs, Git e tantas outras ferramentas—não teria sido possível;

Ao Instituto de Matemática e Estatística da USP, sempre flexível em acomodar minhas responsabilidades profissionais às acadêmicas;

À Fundação Japão, cujas bibliotecas sempre foram uma parte essencial de minha vida intelectual, e cujo programa de bolsas deu-me a oportunidade de pesquisar *in loco*, melhorando consideravelmente a qualidade deste trabalho;

A K., pela inspiração;

A C., pela companhia.

---

<sup>1</sup> <<http://namakajiri.net/nikki>>



*...Tal qual uma epifania, descerá então sobre ti a certeza de que a fantástica singularidade dessas ruas deve-se quase toda à profusão de caracteres chineses e japoneses, pintados em branco, preto, ouro, azul, que tudo ornamentam—até mesmo os umbrais, até mesmo as paredes de papel. Talvez então, por um breve instante, imagines tu o efeito que causaria a substituição de tais caracteres mágicos por letreiros à inglesa, e a simples idéia há de chocar brutalmente tua sensibilidade estética, seja ela qual for; e tornar-te-ás então, assim como eu, um inimigo da Rômaji-Kwai—aquela sociedade fundada com o propósito feio e utilitário de introduzir o uso das letras inglesas para grafar o japonês.*

(Lafcadio Hearn, *Glimpes of Unfamiliar Japan*, 1895)



# Resumo

Existem muitos sistemas de escrita em uso pelo mundo. Quase todos eles são representações dos sons das línguas, compostos por poucas dezenas de símbolos. A escrita japonesa, porém, inclui caracteres chineses (*kanji*), que representam não só os sons mas também os sentidos; e, para isso, precisa empregar milhares de símbolos. A complexidade do sistema de escrita japonês torna-o mais difícil de aprender e de processar mentalmente. Por que então ele continua sendo usado até hoje? Haveria alguma vantagem?

Investigando estas questões, descobrimos que a escrita japonesa permite formas de expressão que não seriam possíveis através da transcrição sonora pura, nem em sistemas de escrita mais simples. Esta observação é importante, não apenas para os estudos japoneses, mas para os estudos da linguagem escrita em geral: o caso japonês demonstra que a escrita não pode ser compreendida como um simples registro visual da fala, mas deve ser estudada como um sistema de acesso à linguagem com características próprias. Neste trabalho, analisamos algumas dessas formas de expressão específicas da escrita, tal como se apresentam no japonês.

**Palavras-chave:** Língua japonesa. Linguística. Sistemas de escrita. *Kanji*. *Hànzì*. *Glossas*. *Furigana*. Língua chinesa. Ideograma. Logografia. Morfografia. Fonografia. Linguística de corpus. Estatística. Técnicas literárias.



# Abstract

There are many writing systems currently in use around the world. For almost all of them, the basic mechanism is using graphical symbols to represent the sounds of language. A few dozen symbols are enough for this purpose. Japanese writing, however, includes Chinese characters (*kanji*), which are related not only to sound but also to meaning; since there are many possible meanings, *kanji* number in the thousands. The complexity of Japanese writing makes it comparatively harder to learn, and harder to process mentally. Why, then, is it still in use? Are there any advantages to such a system?

A closer look show that Japanese writing allows modes of expression which would be impossible in a phonetic transcription of speech, or in simpler writing systems. This is an important datum, not only for Japanese studies, but for the linguistic study of writing itself; the Japanese case clearly shows that writing can't be adequately described as merely a visual representation of speech, but must rather be analyzed as an independent system for accessing language. In this dissertation we discuss, from Japanese examples, some of these expressive techniques which can only be realized in a written medium.

**Keywords:** Japanese language. Linguistics. Writing systems. *Kanji*. *Hànzi*. Glosas. Furigana. Chinese language. Ideogram. Logography. Morphography. Phonography. Corpus linguistics. Statistics. Literary techniques.



# Lista de ilustrações

Figura 1 – Modelo da primazia da fala . . . . .	42
Figura 2 – Comparação entre os modelos da primazia da fala e da autonomia da escrita . . . . .	46
Figura 3 – Variantes da letra ⟨a⟩ . . . . .	58
Figura 4 – Exemplo de rebus (empréstimo homofônico de pictogramas) em português . . . . .	82
Figura 5 – Exemplo de leitura <i>kundoku</i> : “As flores não atrapalham o caminho” .	111
Figura 6 – Distribuição do número de componentes por <i>kanji</i> na amostra, em quartis . . . . .	142
Figura 7 – Frequência de cada <i>kanji</i> na amostra . . . . .	143
Figura 8 – Frequência de cada componente na amostra . . . . .	144
Figura 9 – Os 1% caracteres <i>kanji</i> mais frequentes . . . . .	145
Figura 10 – Histograma da distribuição do melhor índice de consistência fonética para ocorrências de <i>kanji</i> na amostra . . . . .	150
Figura 11 – Histograma da distribuição do melhor índice de consistência fonética para tipos de <i>kanji</i> na amostra . . . . .	151
Figura 12 – Título da obra <i>Hotaru no Haka</i> (“Cemitério dos Vaga-lumes”) . . . . .	157
Figura 13 – Trecho de anúncio publicitário do videogame <i>Shin Megami Tensei IV</i> .	158
Figura 14 – Reanálise morfológica em <i>Ansatsu Kyōshitsu</i> . . . . .	160
Figura 15 – Capa do guia de restaurantes <i>Tōkyō Meibutsu Taberaru Ki</i> 東京名物食べある記, de 1929 . . . . .	161
Figura 16 – Paralelismo no mangá <i>Rurōni Kenshin</i> . . . . .	164
Figura 17 – <i>Furigana</i> duplo com estrangeirismo em <i>Bishōjo Senshi Sailor Moon</i> .	167
Figura 18 – O uso crescente de <i>super</i> スーパー como leitura do <i>kanji</i> 超 . . . . .	169
Figura 19 – Exemplo de paralelismo com língua estrangeira em história em quadrinhos: <i>Deep Submerge!</i> . . . . .	171
Figura 20 – Uso de abreviação em alfabeto romano com glosa <i>furigana</i> . . . . .	173
Figura 21 – <i>Furigana</i> invertido em <i>Green Blood</i> . . . . .	174
Figura 22 – Tipo de caractere como recurso expressivo: <i>Ghost Trick</i> . . . . .	178



# Lista de tabelas

Tabela 1 – Algumas leituras do caractere 雨 . . . . .	29
Tabela 2 – Palavras frequentes cuja grafia inclui o caractere 雨 . . . . .	30
Tabela 3 – Subdivisão da palavra “paraquedismo” em signos . . . . .	52
Tabela 4 – Comparação da escrita portuguesa e variação fonológica . . . . .	56
Tabela 5 – Comparação da escrita portuguesa e pronúncia fonética . . . . .	57
Tabela 6 – Pluralidade na escrita: plurigrafia e plurifonia . . . . .	60
Tabela 7 – Multiplicidade na escrita: pluraridade e polivalência . . . . .	62
Tabela 8 – Morfofonologia do morfema <i>medic-</i> . . . . .	63
Tabela 9 – Usos da letra ⟨h⟩ em português: Heterografia e plurigrafia . . . . .	66
Tabela 10 – Polivalência das letras ⟨e⟩ e ⟨o⟩ e acento gráfico . . . . .	68
Tabela 11 – Exemplos de caracteres chineses . . . . .	70
Tabela 12 – Componentes dos caracteres chineses . . . . .	74
Tabela 13 – Componentes semânticos . . . . .	75
Tabela 14 – Componentes fonéticos . . . . .	77
Tabela 15 – Progressão histórica dos pictogramas chineses ( <i>xiàngxíng</i> ) . . . . .	80
Tabela 16 – Caracteres chineses indicativos ou analógicos ( <i>zhǐ shì</i> ) . . . . .	81
Tabela 17 – Caracteres chineses de associação semântica ( <i>huìyì</i> ) . . . . .	81
Tabela 18 – Caracteres chineses de rebus ou empréstimos homofônicos ( <i>jiǎjiè</i> ) . . . . .	82
Tabela 19 – Caracteres fono-semânticos ( <i>xíng shēng</i> ) . . . . .	83
Tabela 20 – Processo gerativo típico de um caractere fono-semântico . . . . .	84
Tabela 21 – Componentes fono-semânticos . . . . .	85
Tabela 22 – Pronúncias polivalentes eruditas do chinês literário . . . . .	88
Tabela 23 – Exemplos de plurimorfemia no chinês: um caractere denota múltiplos morfemas . . . . .	89
Tabela 24 – Listagem completa de caracteres chineses plurais no dicionário CC-CEDICT . . . . .	90
Tabela 25 – Grafia SA-SB de palavras-borboleta: morfemas dissilábicos cujos caracteres partilham o mesmo componente semântico . . . . .	96
Tabela 26 – O morfema <i>dié</i> . . . . .	97
Tabela 27 – Palavras elásticas no chinês . . . . .	98
Tabela 28 – Tradução de Fenollosa de um verso chinês . . . . .	101
Tabela 29 – Comparação de leituras chinesas e japonesas . . . . .	106
Tabela 30 – Exemplos de múltiplas leituras siníticas atribuídas a um único caractere . . . . .	106
Tabela 31 – Principais fontes dos estratos de leituras <i>on</i> . . . . .	107
Tabela 32 – Leituras <i>kun</i> plurimorfêmicas . . . . .	113
Tabela 33 – Leituras <i>kun</i> compostas ( <i>jukujikun</i> ) . . . . .	114

Tabela 34 – Os <i>hiragana</i> atuais, sua pronúncia e <i>kanji</i> de origem . . . . .	116
Tabela 35 – Os <i>katakana</i> atuais, sua pronúncia e <i>kanji</i> de origem . . . . .	117
Tabela 36 – Relação entre <i>kana</i> , sílabas e moras . . . . .	119
Tabela 37 – Processo de vozeamento <i>daku</i> e o sinal <i>dakuten</i> . . . . .	120
Tabela 38 – Caracteres <i>kana</i> históricos para o som /i/ . . . . .	121
Tabela 39 – <i>Dōon-igi</i> : distinções de nuance ausentes da fala . . . . .	126
Tabela 40 – Alguns morfogramas da escrita cuneiforme . . . . .	128
Tabela 41 – Confusão de componentes fonéticos devido à simplificação gráfica . . . . .	135
Tabela 42 – Contagens de caracteres <i>kanji</i> e leituras na amostra . . . . .	140
Tabela 43 – Relação entre caracteres <i>kanji</i> e número de componentes . . . . .	141
Tabela 44 – Cobertura dos itens mais frequentes sobre o total de ocorrências . . . . .	145
Tabela 45 – Exemplos de séries de componentes . . . . .	146
Tabela 46 – Exemplos de vizinhanças . . . . .	146
Tabela 47 – Exemplos de séries e leituras . . . . .	147
Tabela 48 – Exemplos de consistência fonética . . . . .	147
Tabela 49 – Exemplo de escolha da melhor vizinhança fonética . . . . .	147
Tabela 50 – Índices de consistência fonética das leituras siníticas (on) que ocorreram na amostra . . . . .	149
Tabela 51 – Formas de uso regulares do <i>kanji</i> na escrita japonesa moderna . . . . .	155
Tabela 52 – Grafia das palavras <i>koro</i> “momento, hora” e <i>korosu</i> “assassinar” . . . . .	160
Tabela 53 – Exemplos de estrangeirismos como glosa de <i>kanji</i> propostos por Hashimoto et al. . . . .	168
Tabela 54 – Uso de glosas em alemão em Sasameyuki: “ler com legendas” . . . . .	170
Tabela 55 – Técnicas de escrita de um trecho de <i>Hakuzōsu</i> , de Natsuhiko Kyōgoku. . . . .	181
Tabela 56 – Notação <i>pīnyīn</i> e sua pronúncia . . . . .	215

# Lista de abreviaturas e siglas

## Abreviaturas e siglas no texto

IPA	<i>International Phonetic Alphabet</i> (Alfabeto Fonético Internacional)
jap.	japonês

## Abreviaturas usadas em glosas morfológicas

Sobre abreviações morfológicas, ver o padrão [Leipzig \(2008\)](#).

ADV	Adverbial
ADN	Adnominal, <i>rentaikei</i>
COM	Caso comitativo, companhia
CON	Flexão conectiva, <i>ren'yōkei</i>
COP	Cópula; verbo ou partícula de ligação
CTR	Contador (sufixo de contagem)
DEDU	Modo dedutivo (fato deduzido a partir de evidência indireta)
GEN	Caso genitivo
EXP	Modo experiencial (assevera experiência)
IMP	Modo imperativo
LOC	Caso locativo
NEG	Negativo
NOM	Caso nominativo (sujeito)
NMLZ	Nominalizador, forma nominal
OBJ	Objeto (direto ou indireto)
PRFV	Aspecto perfectivo, ação completa
TOP	Tópico
QUOT	Quotativo, citação



# Lista de símbolos

- [ ] Colchetes denotam transcrição fonética (ver 3.1.1 à página 49).
- // Barras denotam transcrição fonológica (ver 3.1.1 à página 49).
- ⟨ ⟩ Aspas angulares denotam uma forma escrita sendo discutida em si; isto é, onde o que importa são os próprios sinais gráficos, não aquilo que eles normalmente representam. Por exemplo: “Podemos discutir o conceito de amor, a palavra ‘amor’, ou a sequência de quatro letras ⟨amor⟩.”
- \* Asterisco denota a) uma forma histórica reconstruída a partir das modernas, ou b) uma forma que não ocorre na língua.



Os arquivos desta dissertação, incluindo código, dados, e correções/atualizações do texto, encontram-se disponíveis no endereço:

<http://namakajiri.net/letras/mestrado2016>



# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>27</b>
<b>1.1</b>	<b>Sistema de escrita japonês: O pior do mundo?</b>	<b>27</b>
<b>1.2</b>	<b>Objetivos</b>	<b>35</b>
1.2.1	Descrição abrangente da escrita japonesa	35
1.2.2	Consequências para a teoria	36
1.2.3	Investigar possíveis vantagens do sistema de escrita japonês	37
<b>1.3</b>	<b>Metodologia</b>	<b>37</b>
1.3.1	Pesquisa bibliográfica	37
1.3.2	Análise de corpus	38
1.3.3	Coleta de exemplos	38
<b>1.4</b>	<b>Convenções</b>	<b>38</b>
<b>2</b>	<b>FUNDAMENTOS TEÓRICOS: DUAS TEORIAS DA ESCRITA</b>	<b>41</b>
<b>2.1</b>	<b>A primazia da fala</b>	<b>41</b>
<b>2.2</b>	<b>A autonomia da escrita</b>	<b>42</b>
<b>3</b>	<b>SOM E SENTIDO NA ESCRITA: A FONOGRAFIA E A MORFOGRAFIA</b>	<b>47</b>
<b>3.1</b>	<b>A dupla articulação da linguagem: fonemas e morfemas</b>	<b>47</b>
3.1.1	Blocos elementares da linguagem: os fonemas	49
3.1.1.1	Variação fonética condicionada e livre	51
3.1.2	Elementos básicos de sentido: os morfemas	51
3.1.3	A dupla articulação e a escrita	55
<b>3.2</b>	<b>Uma escrita fonográfica: o português</b>	<b>56</b>
3.2.1	Grafos e grafemas	58
3.2.2	Além do princípio fonográfico	59
3.2.2.1	Dígrafos e plurigrafia	59
3.2.2.2	Plurifonia	59
3.2.2.3	Polivalência	60
3.2.2.3.1	Polivalência e profundidade ortográfica	63
3.2.2.3.2	Polivalência e sentido: heterografia	64
3.2.2.4	Símbolos e elementos não-fonográficos	65
3.2.2.4.1	A letra ⟨h⟩	66
3.2.2.4.2	Pontuação, tipografia e sentido	66
3.2.2.4.3	Espaços, sinais diacríticos e heterografia	67
3.2.2.4.4	Algarismos, sinais matemáticos e outros recursos	68
<b>3.3</b>	<b>Uma escrita morfográfica: chinês</b>	<b>69</b>

3.3.1	Contagem dos caracteres chineses . . . . .	73
3.3.2	Estrutura dos caracteres chineses . . . . .	73
3.3.2.1	Contagem dos componentes . . . . .	75
3.3.2.2	Componentes semânticos . . . . .	75
3.3.2.3	Componentes fonéticos . . . . .	77
3.3.2.4	Tipos de caracteres chineses: A classificação tradicional . . . . .	78
3.3.2.5	Componentes de papel duplo: os caracteres fonético-associativos duais . . . . .	85
3.3.2.6	Limitações no papel dos componentes . . . . .	86
3.3.3	Exceções e desvios do princípio morfográfico . . . . .	87
3.3.3.1	Polivalência de leitura na escrita chinesa . . . . .	87
3.3.3.2	Plurimorfemia: um caractere representa uma sequência de morfemas . . . . .	89
3.3.3.3	O mito do monossilabismo? Caracteres sub-morfêmicos e as palavras-borboleta . . . . .	90
3.3.3.4	Caracteres chineses como fonogramas . . . . .	100
3.3.4	O mito do ideograma . . . . .	100
<b>3.4</b>	<b>Fonografia vs. morfografia . . . . .</b>	<b>103</b>
<b>3.5</b>	<b>Uma escrita mista: japonês . . . . .</b>	<b>104</b>
3.5.1	Introdução . . . . .	104
3.5.2	O chinês no japonês: leituras <i>on</i> . . . . .	104
3.5.2.1	Estratos sinóticos: <i>go, kan, tōsō</i> . . . . .	106
3.5.3	Fonografia com o <i>kanji</i> . . . . .	107
3.5.3.1	Terminologia da fonografia em <i>kanji</i> : <i>ateji, kana, Man'yōgana</i> . . . . .	108
3.5.3.2	Fonografia em <i>kanji</i> e a representação do japonês . . . . .	109
3.5.4	A tradução como parte integral da escrita: leituras <i>kun</i> . . . . .	110
3.5.4.1	Tradução além da morfografia: plurimorfemia e os compostos <i>jukujikun</i> . . . . .	113
3.5.5	Fonogramas japoneses: o <i>kana</i> . . . . .	115
3.5.5.1	Irregularidades na fonografia do <i>kana</i> . . . . .	118
3.5.5.2	<i>Kana</i> , sílaba e mora . . . . .	119
3.5.5.3	Diacríticos do <i>kana</i> : <i>dakuten</i> e <i>handakuten</i> . . . . .	119
3.5.5.4	<i>Kana</i> e sons longos . . . . .	120
3.5.5.5	Polivalência do <i>kana</i> histórico: o <i>hentaigana</i> . . . . .	121
3.5.6	A escrita mista: <i>kanji-kana majiribun</i> , flexões e <i>okurigana</i> . . . . .	122
3.5.7	A escrita em paralelo: as glosas <i>furigana</i> . . . . .	123
3.5.8	<i>Kanji</i> , polivalência e ideografia . . . . .	124
3.5.8.1	Mesmo som, outra nuance: a ideografia do <i>dōon-igi</i> . . . . .	125
3.5.9	Sistemas mistos semelhantes ao japonês . . . . .	126
<b>4</b>	<b>MEDINDO O GRAU DE FONETICIDADE DO KANJI . . . . .</b>	<b>131</b>
4.0.10	A proposta de DeFrancis: Seria a morfografia uma fonografia? . . . . .	131
4.0.11	Limitações da análise de DeFrancis . . . . .	132
<b>4.1</b>	<b>O caso do japonês . . . . .</b>	<b>134</b>

4.1.0.1	Funções dos componentes fonéticos . . . . .	135
4.1.0.2	Hipótese . . . . .	136
4.1.1	Detalhes metodológicos . . . . .	136
4.1.1.1	Tipos e ocorrências . . . . .	136
4.1.1.2	Os <i>kanji</i> de uso geral: <i>Jōyō Kanji</i> . . . . .	137
4.1.1.3	O corpus e seu tratamento . . . . .	137
4.1.1.4	Fontes de erros . . . . .	139
4.1.1.5	Outras abordagens . . . . .	139
<b>4.2</b>	<b>Descrição da amostra . . . . .</b>	<b>140</b>
4.2.1	Caracteres <i>kanji</i> e leituras . . . . .	140
4.2.2	Componentes . . . . .	141
4.2.3	Frequências . . . . .	142
<b>4.3</b>	<b>Análise do corpus . . . . .</b>	<b>145</b>
4.3.1	Definições . . . . .	145
4.3.2	Pressupostos da análise . . . . .	148
4.3.3	Medindo a consistência das pistas fonéticas dos <i>kanji</i> . . . . .	149
<b>4.4</b>	<b>Discussão e perspectivas . . . . .</b>	<b>151</b>
<b>5</b>	<b>TÉCNICAS DE EXPRESSÃO CRIATIVA USANDO KANJI . . . . .</b>	<b>155</b>
<b>5.1</b>	<b>Reanálise morfológica . . . . .</b>	<b>155</b>
5.1.1	Reanálise semântica . . . . .	156
5.1.2	Reanálise lúdica . . . . .	162
5.1.3	Reanálise de estrangeirismo . . . . .	163
<b>5.2</b>	<b>Paralelismo . . . . .</b>	<b>163</b>
5.2.1	Com palavras japonesas . . . . .	163
5.2.2	Com estrangeirismos . . . . .	165
5.2.3	Com outros tipos de caracteres . . . . .	172
<b>5.3</b>	<b>Logografia e criatividade em nomes próprios . . . . .</b>	<b>175</b>
<b>5.4</b>	<b>Outras técnicas gráficas . . . . .</b>	<b>176</b>
5.4.1	O tipo de caractere como recurso expressivo . . . . .	176
5.4.2	Peso . . . . .	179
5.4.3	Estilo . . . . .	180
<b>5.5</b>	<b>A escrita <i>kanji</i> na concepção popular . . . . .</b>	<b>184</b>
<b>6</b>	<b>CONCLUSÕES . . . . .</b>	<b>187</b>
	<b>Referências . . . . .</b>	<b>189</b>

<b>APÊNDICES</b>	<b>203</b>
<b>APÊNDICE A – GLOSSÁRIO DE TERMOS TÉCNICOS . . . . .</b>	<b>205</b>
<b>APÊNDICE B – GLOSSÁRIO DE TERMOS ESTRANGEIROS . . . . .</b>	<b>213</b>
<b>APÊNDICE C – A TRANSCRIÇÃO DO MANDARIM PADRÃO <i>PĪNYĪN</i> . . . . .</b>	<b>215</b>

# 1 Introdução

## 1.1 Sistema de escrita japonês: O pior do mundo?

O sistema de escrita japonês constitui, talvez, a mais complicada forma de escrever já imaginada. (FISCHER, 2004, p. 167)

Geralmente considerado o sistema de escrita mais complicado do mundo [...] (ROBINSON, 2007, p. 205)

É necessário que eu diga algo sobre a escrita japonesa, amplamente considerada a mais complexa atualmente em uso. (SPROAT, 2000, p. 132)

Em mãos japonesas, os caracteres chineses transformaram-se no que é com frequência chamado de o sistema de escrita mais intrincado, o mais complicado já usado por uma população de tamanho razoável. [...] Nem mesmo o sistema original chinês se compara à adaptação japonesa, no que diz respeito à complexidade e aparente impraticabilidade. (COULMAS, 1989, p. 122)

O sistema de escrita japonês moderno ainda retém a distinção dúbia de ser o mais complexo em uso no mundo moderno. (SEELEY, 1984, p. 213)

Foi apenas a partir de 1950 que todas as crianças japonesas em idade escolar passaram a ter que suportar a plena carga do que é considerado, por consenso geral, o sistema de escrita mais oneroso atualmente em uso. (UNGER, 2004, p. 11)

Os japoneses terminaram por ficar com um dos piores sistemas de escrita jamais criados. (DEFRANCIS, 1989, p. 138)

A ortografia japonesa partilha com a inglesa a distinção de ser a pior da sua categoria, exceto que ao invés de ser a pior da categoria boa, ela é a pior da categoria ruim. [...] Como os fabricantes da Lexus LS-400<sup>1</sup> e dos melhores microeletrônicos do mundo conseguem viver com um sistema de escrita como esse, é um fato que está além da possibilidade de compreensão. (HANNAS, 1997, p. 26–28)

Temos de hesitar ao tentar encontrar um epíteto que descreva um sistema de escrita tão complexo, que ele precisa do apoio de um outro sistema de escrita apenas para explicá-lo.<sup>2</sup> A escrita japonesa, sem dúvida, prova ser uma área de estudos deveras fascinante; mas, enquanto ferramenta prática, ela certamente não tem inferiores. (SANSOM, 1928, p. 44)

<sup>1</sup> Modelo de luxo dos automóveis Toyota.

<sup>2</sup> Sansom refere-se às glosas *furigana*; ver seção 3.5.7 à página 123.

A escrita chinesa é a pior escrita do mundo, exceto por uma, e esta é derivada daquela. Falo, é claro, da escrita sobre a qual Sir George Sansom disse não haver inferiores: a monumental escultura de lixo que os japoneses erigiram através da bricolagem impiedosa de textos chineses... (CIKOSKI, 2011, p. x).

Algumas das citações acima talvez soem como julgamentos etnocêntricos e desinformados da era colonial, mas não são. Todos os trechos são de trabalhos técnicos de linguística, de autoria de renomados especialistas—profundos conhecedores das várias escritas do mundo, e da japonesa em particular. Mais para o fim da lista, o tom ácido, indisfarçado, destoa do restante das obras dos mesmos autores, quase todo composto de análises objetivas, sem julgamento de valor. [Gottlieb \(1995\)](#) também comentou sobre como, no Japão, os debates intelectuais sobre a questão parecem ser particularmente emotivos (p. 38). Por que a escrita japonesa provoca opiniões tão fortes?<sup>3</sup>

O ponto polêmico é o uso dos *kanji*—literalmente “caracteres chineses”. Originalmente criados para representar o chinês literário, esses caracteres foram adaptados para o japonês progressivamente a partir do século VIII ([SEELEY, 1991](#)). Ao contrário da maioria das escritas atualmente em uso no mundo, os *kanji* não representam somente o som, mas também o sentido das palavras. Como há mais sentidos do que sons, o número de caracteres chineses alcança as dezenas de milhares. Além disso, a adaptação para o japonês complexificou seu mecanismo de leitura. A representação da língua japonesa com *kanji* resulta nas seguintes desvantagens ([SMITH, 1996](#), p. 214):

- Para poder ler textos em japonês, é preciso memorizar milhares de caracteres. O sistema de ensino japonês ensina 2136 caracteres denominado de “uso geral” (*Jōyō*); para participação na vida letrada adulta, este conjunto é o mínimo a ser memorizado ([MEXT, 2010](#)). Na prática muitos outros se fazem necessários, como caracteres especiais usados em nomes próprios (pessoais ou topônimos).
- Além disso, cada caractere pode possuir várias leituras, multiplicando ainda mais a carga na memória. Por exemplo, 山 “montanha” pode ser lido como *yama* ou *san*, e, no caso extremo, 生 “vida” possui 12 leituras distintas—sem contar leituras especiais usadas em nomes próprios. O padrão *Jōyō* especifica aproximadamente 4400 leituras, de acordo com nossa contagem.<sup>4</sup>
- Além disso, há grande quantidade de caracteres homófonos—ou seja, uma mesma sílaba pode ser representada por inúmeros caracteres distintos. A sílaba *san* pode ser representada por qualquer um de 三 山 参 栈 蚕 惨 産 傘 散 算 酸 賛, depen-

<sup>3</sup> Esta amostra de citações foi ampliada a partir da lista de [Joyce \(2012\)](#).

<sup>4</sup> Ver programa *JōyōDB* nos arquivos dessa dissertação.

dendo do sentido.<sup>5</sup>

- Além disso, não há consistência sobre qual unidade linguística os caracteres representam: eles podem denotar apenas uma sílaba, como *san*, ou até cinco, como 政 *matsuri-goto* “governo”; podem representar menos do que uma palavra, como 大人 *otona* “adulto”, palavra indivisível escrita com dois caracteres; ou uma composição inteira, como 志 *kokoro-zashi*, literalmente “direção da mente” = “intenção.
- E, com todas essas possibilidades, torna-se difícil prever qual a leitura correta em cada situação.<sup>6</sup>

Historicamente, escritas de funcionamento semelhante ao dos caracteres chineses incluem os hieróglifos egípcios, o cuneiforme sumério e a escrita maia. Adaptações comparáveis à escrita japonesa incluem antigas formas de escrita acádia (babilônia), coreana e vietnamita. A invenção da escrita simples, baseada no som, é mais recente, e foi adotada por quase todas as culturas, incluindo Coreia e Vietnã. Por que então a China, o Japão e Taiwan insistem em usar formas mais antigas e complexas de escrita?

Para entender melhor o que os linguistas citados consideram “complexidade”, consideremos o caractere “chuva”, 雨 (em negrito nos exemplos):

**Tabela 1 – Algumas leituras do caractere 雨**

Palavra	Escrita japonesa	Leitura de 雨	Tradução
<i>Ame</i>	<b>雨</b>	<i>ame</i>	“chuva”
<i>Ama-dare</i>	<b>雨だれ</b>	<i>ama</i>	“gotas de chuva”
<i>Haru-same</i>	<b>春雨</b>	<i>same</i>	“chuva de primavera”
<i>Bai-u</i>	<b>梅雨</b>	<i>u</i>	“estação das chuvas”
<i>Tsuyu</i>	<b>梅雨</b>	Em conjunto com 梅, <i>tsuyu</i>	“estação das chuvas”

Nos três primeiros exemplos, a palavra japonesa *ame* altera-se para *ama-* ou *same*; mas essa variação não aparece na escrita. As três variações são todas grafadas 雨. Para saber como ler as formas 雨 *ame*, 雨だれ *amadare* e 春雨 *harusame*, é necessário memorizá-las uma por uma.

O quarto exemplo, *baiu*, é um empréstimo do chinês.<sup>7</sup> A palavra significa literalmente a “chuva” (*u*) das “ameixeiras” (*bai*), e se refere a um período chuvoso no meio

<sup>5</sup> 三 “três”, 山 “montanha”, 参 “participar”, 栈 “suporte de madeira”, 蚕 “bicho-da-seda”, 惨 “desastre”, 産 “produzir”, 傘 “guarda-chuva”, 散 “dispersar”, 算 “calcular”, 酸 “ácido” e 贊 “aprovar”, todos *san*. Se incluirmos caracteres e pronúncias raras, do tipo listados em um dicionário, o número de caracteres para *san* chega a 132 (BREEN et al., 2016). Essas pronúncias são derivadas do chinês clássico, onde toda sílaba carrega um sentido; cf. seção 3.3.3.3.

<sup>6</sup> Para um estudo empírico sobre o impacto dessa complexidade no processamento mental, ver Kaya-moto, Yamada e Takashima (1998).

<sup>7</sup> A pronúncia em chinês medieval era algo como \**muai-yu*<sup>B</sup> (SCHUESSLER, 2009, pp. 377; 587); notação fonética do autor.

do verão, quando as ameixeiras vicejam exuberantes. Note que o empréstimo chinês *u*, “chuva”, é escrito de forma idêntica às palavras nativas japonesas *ame*, *ama-*, e *-same*. Ou seja, todas estas quatro palavras são leituras possíveis do caractere 雨.

O quinto exemplo, *tsuyu*, é o mais complexo. Observe, em primeiro lugar, que a grafia das palavras japonesas *baiu* e *tsuyu* é idêntica: 梅雨. Apenas o contexto e experiência permitem presumir qual das duas leituras o escritor tinha em mente. Essas palavras formam um par de quase-sinônimos: *baiu*, de origem chinesa, é mais literária, enquanto que o equivalente nativo, *tsuyu*, soa mais natural.<sup>8</sup>

Ainda sobre o quinto exemplo, em 梅雨 *tsuyu*, o caractere 雨 não tem leitura individual; a palavra precisa ser lida como um todo indivisível. Isso acontece porque a grafia chinesa da palavra, 梅雨, com dois caracteres, foi adotada na íntegra para representar sua tradução em japonês, *tsuyu*. A tradução não é literal; *tsuyu* se refere ao mesmo período chuvoso, mas é um termo indivisível, que não se decompõe em “chuva das ameixeiras”.<sup>9</sup> Não é o caso que um caractere represente *tsu* e outro *yu*; é o par de caracteres 梅雨, tomado em conjunto, que representa a palavra inteira *tsuyu* (*jukujikun*; ver seção 3.5.4.1 à página 113).

As seguintes palavras incluem o caractere 雨 na grafia, e estão marcadas no dicionário EDICT como “frequentes” (BREEN et al., 2015):

**Tabela 2 – Palavras frequentes cuja grafia inclui o caractere 雨**

Leitura	Palavras	Quantidade
<i>ame</i>	<i>ame</i> , <i>ko-ame</i> , <i>oo-ame</i> , <i>ame-furi</i> , <i>naga-ame</i> , <i>niwaka-ame</i>	6
<i>ama-</i>	<i>ama-gumo</i> , <i>ama-gu</i> , <i>ama-do</i> , <i>ama-yadori</i> , <i>ama-mori</i> , <i>ama-mizu</i> , <i>ama-dare</i>	7
<i>-same</i>	<i>kiri-same</i> , <i>ko-same</i>	2
<i>u</i>	<i>u-ten</i> , <i>u-ryō</i> , <i>kō-u</i> , <i>gō-u</i> , <i>ji-u</i> , <i>fū-u</i> , <i>rai-u</i> , <i>bai-u</i> , <i>u-ki</i>	9
Outras	<i>tsuyu</i> , <i>shigure</i>	2

Todas as 26 ortografias precisam ser aprendidas caso a caso.<sup>10</sup>

E tudo o que foi discutido até agora são as leituras de um único caractere. Como mencionado acima, para funcionar como cidadão letrado na sociedade japonesa, todo estudante precisa aprender na escola cerca de 4407 leituras de 2136 caracteres—processo completado apenas na conclusão do equivalente ao nosso ensino médio (MEXT, 2010).

<sup>8</sup> Isso é comparável à forma que o português distingue empréstimos gregos eruditos, e.g. *geológico* ou *heliocêntrico*, de palavras nativas latinas, como *terrestre* e *solar*.

<sup>9</sup> A etimologia mais provável liga a “orvalho”, também *tsuyu* (LOOKWISE, 2016).

<sup>10</sup> Ao todo, o dicionário EDICT inclui atualmente 279 entradas com o caractere 雨. Porém, não está claro quantas entradas são previsíveis a partir de outras, já que o dicionário inclui palavras compostas e expressões idiomáticas. O número que apresentamos é uma estimativa conservadora. Das palavras que foram omitidas por baixa frequência, destacamos 春雨 *haru-same* “chuva de primavera” e 五月雨 *samidare* “chuvas de maio”.

Começamos a entender o porquê das opiniões negativas sobre a escrita japonesa.

Há ainda mais uma crítica ao uso do *kanji*: a de que todo esse esforço de memorização seria desnecessário, perfeitamente dispensável (MATSUNAGA, 1994, pp. 122–123). A escrita portuguesa, por exemplo, usa apenas 26 caracteres para representar qualquer palavra. Ela consegue isso porque as letras representam os sons da língua, que são poucos—menos de 40. Não seria mais racional representar o japonês também com uma escrita assim? E mais, o japonês já possui uma disponível: o *kana*, que com 46 símbolos consegue representar qualquer palavra japonesa de forma simples e direta, pelo som. Mas, na grafia atualmente em uso, os *kana* intercalam-se com *kanji* no mesmo texto. Eles são usados até mesmo em paralelo: quando um escritor estima que um *kanji* será muito difícil de ler, ele especifica sua leitura através de glosas (anotações paralelas) em *kana*.<sup>11</sup> Mas então por que se dar ao trabalho de usar *kanji*, em primeiro lugar? Se já temos em mãos o *kana*, que é mais simples, por que não simplesmente escrever tudo em *kana*?

Historicamente, a Coreia e o Vietnã usavam sistemas mistos semelhantes ao japonês. Ambos os países abandonaram os caracteres chineses com sucesso (COULMAS, 1989, cap. 7). Os críticos que vimos acima defendem que o Japão deveria fazer o mesmo, especulando que isso traria diversos benefícios sociais. Por exemplo, Unger e Matsunaga acreditam que existem diferenças no nível de aquisição de *kanji*, o que implicaria que o uso de *kanji* estaria aumentando a desigualdade social (MATSUNAGA, 1994, p. 139; UNGER, 1996). Unger argumenta ainda que essa desvantagem na escola seria uma causa do alto índice de suicídio estudantil (UNGER, 1987, p. 94–95).<sup>12</sup> Unger e DeFrancis afirmam que a escrita japonesa seria difícil de processar digitalmente, o que diminuiria a competitividade do Japão no mercado informatizado (DEFRANCIS, 1989, pp. 266–268; UNGER, 1987). Hannas afirma que a didática do *kanji*, baseada na memorização, reforçaria uma cultura de conformismo, atrasando a inovação científica (HANNAS, 1997). Para Pullum, os caracteres chineses são um “pesado fardo” e um “grande freio no desenvolvimento da educação, comércio e ciência” (PULLUM, 2016). Tudo isso, argumentam os linguistas da escola crítica, iria melhorar drasticamente se os caracteres chineses fossem simplesmente abandonados.

No Japão, já no período Meiji (1868–1912) aparecem propostas de simplificar a escrita—seja adotando o alfabeto romano, seja a grafia nativa *kana*. Ainda hoje existem grupos que defendem a extinção dos *kanji*, como o Nippon-no-Rômazi-Sya<sup>13</sup> e o Kanamojikai.<sup>14</sup> Porém, tais grupos sempre foram minoria; o público em geral não parece interessado em simplificar a escrita (UNGER, 1996, p. 127). Em uma pesquisa de opi-

<sup>11</sup> Cf. seção 3.5.7 à página 123.

<sup>12</sup> Considere, porém, que a Coreia possui uma escrita considerada amplamente como simples e lógica, e contudo o índice de suicídio estudantil é um dos maiores do mundo, bem à frente do Japão (THE JAPAN TIMES, 2013).

<sup>13</sup> “Sociedade Japonesa de Caracteres Romanos”; <<http://www.age.ne.jp/x/nrs/>> .

<sup>14</sup> “Associação dos Caracteres Kana”; <<http://www9.ocn.ne.jp/~kanamozi/>> .

nião do governo japonês em 2010, 72,4% dos entrevistados julgaram indispensáveis os caracteres *kanji*, e uma média de 91,3% afirmou não ter dificuldades em decodificá-los ao ler jornais, revistas e páginas da Internet (BUNKACHŌ, 2010). Quando a lista atual de caracteres de uso geral (Jōyō Kanji) estava em desenvolvimento, em 1977, uma pesquisa junto ao público revelou uma tendência a aprovar toda adição de caracteres extras, e rejeitar a remoção de caracteres (GOTTLIEB, 1995, p. 191). A tendência atual na escrita japonesa tem sido em aumentar o número de *kanji*, ao invés de diminuí-lo; a lista de caracteres do sistema escolar passou de 1850 (MEXT, 1946) para 1945 (MEXT, 1981) para 2136 (MEXT, 2010); e, ao contrário do que Unger (1987) e Hannas (1997) previram, a difusão dos computadores e celulares facilitou o uso de um número ainda maior de *kanji*, ao invés de limitá-lo (MATSUNAGA, 1994, p. 15; TABATA, 2015; KESS; MIYAMOTO, 1999, p. 208).<sup>15</sup> E se por um lado o Japão tem “a escrita mais complexa do mundo”, por outro o país mostra uma taxa de letramento altíssima, chegando aos 99% da população (CIA, 2015).

Como vimos, a escrita japonesa desperta em alguns especialistas uma retórica especialmente ácida, agressiva. Porém, em outros casos a reação é o inverso: igualmente emocional, mas apaixonada, elegíaca. Considere o trecho de Lafcadio Hearn citado na epígrafe deste trabalho (HEARN, 1895, p. 4):

...Tal qual uma epifania, descerá então sobre ti a certeza de que a fantástica singularidade dessas ruas deve-se quase toda à profusão de caracteres chineses e japoneses, pintados em branco, preto, ouro, azul, que tudo ornamentam—até mesmo os umbrais, até mesmo as paredes de papel. Talvez então, por um breve instante, imagines tu o efeito que causaria a substituição de tais caracteres mágicos por letreiros à inglesa, e a simples idéia há de chocar brutalmente tua sensibilidade estética, seja ela qual for; e tornar-te-ás então, assim como eu, um inimigo da *Rômaji-Kwai*—aquela sociedade fundada com o propósito feio e utilitário de introduzir o uso das letras inglesas para grafar o japonês.

Na citação da página 27, Sansom se mostra chocado com a combinação de *kanji* e *kana* do japonês: um sistema tão complicado que precisa de outro para explicá-lo. Ariga, porém, aprecia o potencial literário que só existe através dessa combinação (ARIGA, 1989, p. 335):

Tal *marginalia* pode ter sido originalmente acessória; porém, inúmeros estilos textuais empregaram-na para causar os mais variados efeitos lite-

<sup>15</sup> Embora a habilidade de escrever os caracteres à mão pareça estar sendo erodida (MAIR, 2016a). A memória ativa ou de recuperação (*recall*) sempre é mais limitada que a memória passiva ou de reconhecimento; por exemplo, nós entendemos mais palavras do que empregamos. Para escrever caracteres sem consultar referências, é preciso usar memória ativa; conseqüentemente, o número de caracteres que um indivíduo consegue ler sempre é maior do que o que consegue escrever à mão. A escrita digital aproveita-se disso, listando para o escritor sugestões de caracteres para reconhecimento. No computador ou celular, o escritor de *kanji* só precisa ser capaz de reconhecê-los passivamente. Por isso, o número de caracteres em uso aumentou, mas a habilidade de escrevê-los sem referência caiu em desuso e enfraqueceu.

rários. Nos três exemplos de textos *gesaku* examinados acima, as glosas *rubi* [=furigana] são vitais na criação de imagens compostas e de espaço semântico [...] Somente examinando sua função é que entendemos o intrincado jogo polifônico e polissêmico que se encontra na literatura japonesa.

Também [Backhouse \(1984\)](#) observou que:

A presença de uma pluralidade de subsistemas permite flexibilidade ortográfica em uma escala tal que seria inimaginável nos sistemas de escrita mais conhecidos. [...] esta flexibilidade é amplamente explorada, em grau diferente dependendo do gênero textual; e as implicações estilísticas de tal variação merecem mais estudo do que têm recebido.

Ernest Fenollosa, em um ensaio cuja correção linguística viria a ser severamente questionada ([KENNEDY, 1958](#)), mas que foi grande influência em poetas como Ezra Pound e Haroldo de Campos ([CAMPOS; DANTAS, 1994](#)), elogia o potencial poético da escrita chinesa porque os caracteres chineses são *motivados*—ou seja, ao contrário do som das palavras, que é arbitrário,<sup>16</sup> a forma visual (pelo menos em alguns caracteres frequentes) traz uma relação com o que significam ([FENOLLOSA et al., 1936](#)). Por exemplo, o caractere “chuva” 雨 é baseado na ilustração de uma nuvem contendo quatro gotas d’água. Para Fenollosa (pp. 6;55),

A notação chinesa vai muito além de símbolos arbitrários; ela é baseada em uma ilustração vívida e condensada das operações da natureza. Na variável algébrica e na palavra falada não há conexão natural entre objeto e signo; tudo depende de pura convenção; mas o método chinês segue a sugestão natural. [...] Estou convencido que a linguagem escrita chinesa não apenas absorveu a substância da natureza e eregiu com ela um segundo mundo de metáforas, mas que ela, por sua própria visibilidade pictórica, pôde reter a criatividade poética original com muito mais vigor e vivacidade que qualquer linguagem [=sistema de escrita] fonética.

É verdade que muitas das associações pictóricas que Fenollosa faz são imaginativas, e não correspondem à lógica original dos caracteres (cf. seção 3.3.2.4 à página 78). Porém, pondera Haroldo de Campos (op.cit., p. 51),

A mente de um poeta não resistiria a essa viquiana “fábula em miniatura” que parece pulsar na justaposição do “radical” e do “fonograma” [os elementos do caractere].

<sup>16</sup> A “arbitrariedade do signo” é uma doutrina clássica da linguística, mas estudos recentes nos campos do simbolismo sonoro e línguas de sinais têm relativizado a questão. Há agora bastante evidência estatística que pelo menos parte dos sons da fala está relacionado aos sentidos de forma icônica; as línguas preferem, por exemplo, vogais posteriores (mais ressonantes) para palavras associadas à grandeza ou magnitude. Ver e.g. [Perniss, Thompson e Vigliocco \(2010\)](#); [Berlin \(1992, cap. 6\)](#), et al.

Essa vivacidade dos caracteres chineses teve influência no pensamento dos próprios chineses e japoneses (ver por exemplo nota sobre o *Shuōwén Jiězì* à página 78). Ela também influenciou diversas correntes artísticas pelo mundo, como a poesia imagista de Pound, a poesia concreta brasileira, e criadores de outras áreas como o cineasta Sergei Eisenstein (EISENSTEIN, 1949, pp. 37;42):

O japonês domina uma quantidade aparentemente ilimitada de hieróglifos. [...] Convivem com estes uma série de alfabetos fonéticos à européia: o *Man'yōgana*, o *hiragana* e outros. Contudo, o japonês escreve em todas as letras, empregando as duas formas de escrita de uma só vez!

[...] A imagem de “água” junto à de “olho” quer dizer “chorar” [泪]; enquanto a de “orelha” sob o desenho de um “portal” representa o verbo “ouvir” [聞]. [...] Mas isto nada mais é que é—montagem! Sim, exatamente o que fazemos no cinema, combinando cenas que são denotativas, de significado único, de conteúdo neutro, em contextos e sequências intelectualizadas. Isso é o meio e método inevitável de toda exposição cinematográfica; e, em forma condensada e purificada, o ponto de partida do “cinema cult”—do cinema que busca o máximo de laconismo na representação visual de conceitos abstratos. E nós saudamos o método do tão lamentado Cangjie [o criador mítico dos caracteres chineses] como um primeiro passo nesse caminho.<sup>17</sup>

A visão que classifica os caracteres chineses como uma escrita inferior insere-a uma linha evolutiva universal, que parte do desenho, passa pela escrita pelo sentido, evolui para a escrita pelo som, e culmina no alfabeto segmental, tido como ápice da eficiência (DEFRANCIS, 1984; BOLTZ, 1994). Tomando esse pressuposto, é preciso explicar por que os chineses e japoneses pararam de “evoluir” e estacionaram em uma etapa “primitiva”. O linguista Stephen Dodson,<sup>18</sup> em discussão com o autor desta dissertação, levantou a seguinte imagem (MAIR, 2016b):

É como se todos fossem obrigados a gastar anos aprendendo a fazer tijolos de cerâmica e erigir com eles diversas formas de construções, e aqueles que já completaram o aprendizado ficassem indignados com a ideia que as gerações futuras deveriam ser poupadas de continuar o ritual [...] Minha posição é que seria um progresso se todos não fossem obrigados a desperdiçar anos de sua vida aprendendo a fazer tijolos.

Similarmente, uma pesquisadora egípcia da língua japonesa ponderou (comunicação oral): “Nós paramos de escrever com hieróglifos no tempo dos faraós! Por que afinal os japoneses ainda usam isso até hoje?”

A crítica dessa noção evolutiva das escritas é um dos argumentos principais do mais influente livro de Derrida, *De la grammatologie* (p. 91):

<sup>17</sup> Para mais detalhes sobre a juxtaposição de sentidos que Eisenstein denominou “montagem”, ver seções 3.3.2.4 (p. 78) e 3.3.4 (p. 100).

<sup>18</sup> Autor do sítio <<http://languagehat.com>>.

Se a tarefa é analisar realmente aquilo que, sob o nome de “escrita”, distingue muito mais do que técnicas de notação, não é importante abandonar, dentre outros pressupostos etnocêntricos, essa espécie de monogeneticismo gráfico que transforma todas as diferenças em divergência, atraso, acidente, desvio?

Para Derrida, a escrita chinesa desenvolveu-se “fora do logocentrismo”—conceito filosófico do autor que propõe uma falácia no pensamento greco-europeu: a noção de “significado” como uma presença independente, transcendental, que existe à parte da realidade concreta (BOHM; STATEN; CHOW, 2001).

Os caracteres chineses exercem fascínio, atração, deleite estético. Muitos japoneses expressam orgulho nacionalista dos *kanji*, mesmo sabendo que o sistema não é nativo (MATSUNAGA, 1994, p. 122). Estrangeiros que sequer entendem como funcionam os caracteres investem dinheiro para tê-los decorando as casas, as roupas, até mesmo o próprio corpo (TANG, 2015). Unger, crítico, comentou sobre essa tendência com visível desdém (UNGER, 2004, p. xvi):

A sedução do *kanji* tem também um aspecto estético que muitas vezes leva a uma paixão pelos gostos da caligrafia do Extremo Oriente. [...] Em casos extremos, o aficionado começa a perceber um padrão grandioso subjacente a todos os caracteres, ignorado até pelos próprios asiáticos. Tal qual um jogador de xadrez memorizando aberturas, ele grava cada novo caractere na memória como se estivesse tomando um esteroide para o cérebro, ou armazenando uma nova pérola de sabedoria em alguma caixinha de joias mentais. Cedo ou tarde, quase todo estudante de uma língua do Extremo Oriente cai presa de tais sentimentos, ou conhece um colega que caiu. Este livro é para eles—não para desencorajá-los ou diminuir seu apreço pela grande floresta do *kanji*, mas sim para qualificá-lo, colocando a floresta contra um paisagem mais ampla e ensolarada. [...] Nada aqui está além do alcance do leitor leigo—especialmente aquele que está contemplando alguns meses, ou quiçá toda uma vida, de devoção monástica a um rosário de caracteres chineses.

Mas Unger não se detém no porquê dessa estranha atração. O que há na escrita *kanji* que motiva opiniões tão fortes e tão contrárias? Por que os japoneses insistem em continuar usando um sistema tão complicado? As desvantagens dessa escrita estão claras, mas haveria alguma vantagem—algo que só ela pode fazer?

## 1.2 Objetivos

### 1.2.1 Sintetizar uma descrição abrangente do sistema de escrita japonês

Dada a posição particularmente polêmica da escrita japonesa, julgamos frutífero fazer uma revisão ampla da literatura acadêmica, buscando integrar, em um modelo abrangente, os argumentos e os dados levantados em diversas áreas. Nossa síntese

busca explicar como o sistema funciona, não só em seus mecanismos básicos e regulares, mas também as técnicas incomuns ou excepcionais. Com isso, buscamos tornar claro que pontos o sistema de escrita japonês tem em comum com os outros, e em que difere.

Lurie (2006) observou que a crítica da escrita japonesa está associada a um discurso que contrapõe a linguística, enquanto disciplina institucional, a outras áreas, como a filosofia da língua e a crítica literária. Ou seja, muitos dos debates têm um caráter disciplinar, sendo enquadrados por linguistas como defesa da verdade “científica” contra ataques “de fora”. Por isso, embora o foco desta dissertação seja a linguística, será produtivo considerar também trabalhos “de fora” (incluindo filosofia, teoria literária, estudos culturais, psicologia, neurologia etc.), pois cada perspectiva esclarece aspectos diferentes da escrita japonesa; a mesma complexidade que, para um linguista, parece uma representação “ineficiente”, pode ser, para um literato, “riqueza de expressão”. Esperamos que esta revisão abrangente da literatura ofereça uma modesta contribuição ao diálogo interdisciplinar.

### 1.2.2 Explorar as implicações da escrita japonesa para a teoria linguística

Para Joyce (2012), a escrita japonesa é de grande interesse teórico e merece “atenção especial” precisamente por ser incomumente complexa e polivalente. Um modelo científico precisa tentar descrever todos os dados, inclusive os casos extremos e atípicos. Se um modelo teórico da linguagem escrita não dá conta de explicar os fatos da escrita japonesa, isso significa que esse modelo é demasiadamente limitado. Na hierarquia de Chomsky (1969), uma teoria científica da linguagem precisa buscar ser “descritiva” e “explicativa”: com o que quer dizer, precisa descrever as formas que o falante produz e deixa de produzir (“adequação descritiva”), e também os mecanismos que permitem ao falante adquirir essa capacidade (“adequação explicativa”; pp. 18–27). Aplicando os mesmos princípios à teoria da escrita, conclui-se que uma boa teoria precisa, no mínimo, ser capaz de descrever todas as formas de escrita empregadas pelos seres humanos. Sem essa base, não há chance de alcançar o nível explicativo, que é o que possui maior poder preditivo enquanto teoria científica (p. 26).

Nesta dissertação, comparamos duas propostas teóricas opostas: a de que a escrita seria essencialmente uma transcrição da língua falada; e a de que a escrita seria um sistema autônomo, isto é, uma aplicação independente dos princípios da linguagem (ver capítulo 2 à página 41). Argumentamos que as particularidades da escrita japonesa são descritas mais adequadamente pelo segundo modelo, e apresentamos exemplos nos capítulos seguintes.

### 1.2.3 Investigar possíveis vantagens do sistema de escrita japonês

As desvantagens da escrita japonesa estão bem discutidas na literatura linguística (DEFRANCIS, 1989; UNGER, 2004; ERBAUGH, 2002). O que talvez não esteja tão claro são os motivos de seu uso continuado. Segundo o argumento de Unger (1996), a escrita japonesa continua em uso sobretudo por inércia, somada ao conservadorismo político. No entanto, como discutido na seção 1.1, a opinião pública japonesa é bastante favorável à continuidade do sistema, e muitos observadores expressam admiração ou fascínio. A permanência desta escrita deve-se puramente a “apegos sociais, culturais e emocionais” (MATSUNAGA, 1994, p. 138–141), ou haveriam vantagens concretas? Se a escrita fosse substituída por uma mais simples, algo seria perdido? Para esclarecer estas questões, analisamos a escrita japonesa quantitativamente (capítulo 4), e elencamos exemplos de formas de expressão que só são possíveis com os mecanismos próprios a ela (5).

## 1.3 Metodologia

Este trabalho emprega três métodos:

- Pesquisa bibliográfica para síntese teórica (capítulo 3);
- Análise de corpus (capítulo 4);
- E argumentação por meio de contra-exemplos (capítulo 5).

### 1.3.1 Pesquisa bibliográfica

A metodologia principal deste trabalho é a síntese teórica a partir de pesquisa bibliográfica. Isso é consistente com o objetivo de uma dissertação de mestrado, *viz.*, investigar em profundidade a literatura já existente sobre um tema (SIBI/USP, 2009, p. 15). Porém, mais do que simplesmente descrever as teorias existentes, buscamos integrá-las em um todo coerente, investido de senso crítico, integrado com os resultados de estudos experimentais. Esta síntese compõe o capítulo 3.

A escrita japonesa é considerada híbrida ou mista, abarcando dois princípios: a representação pelo som e pelo sentido. Pra ilustrar, discutimos primeiro o alfabeto do português, como exemplo da escrita pelo som, e a escrita chinesa, como exemplo da escrita pelo sentido. A análise da escrita japonesa é então apresentada como uma combinação de técnicas presentes em ambas.

### 1.3.2 Análise de corpus

Uma proposta sob a qual nos detemos mais tempo é a afirmação de DeFrancis de que a escrita chinesa seria “66% fonética” (DEFRAncIS, 1984, p. 109)—dado a partir do qual o autor conclui que todas as escritas são fonéticas. Mas e no caso japonês? A fim de estimar uma medida semelhante para o *kanji*, empregamos a análise estatística de corpus, apresentada no capítulo 4. O corpus escolhido foi a Wikipedia japonesa, por sua acessibilidade, diversidade de tópicos, e magnitude (mais de 2 bilhões de palavras, na edição coletada). As leituras dos caracteres *kanji* desse corpus, porém, precisaram ser inferidas automaticamente. Usamos para isso a ferramenta Mecab (KUDŌ, 2013), cuja taxa de acertos é estimada como superior a 96% (MORI, 2011), e que teve os melhores resultados em nossos testes.<sup>19</sup> A análise em si foi feita com as ferramentas Python (PYTHON FOUNDATION, 2015) e R (R CORE TEAM, 2015). Usamos também informações sobre *kanji* dos projetos Kanjidic (BREEN et al., 2016) e KANJIVG (APEL et al., 2013).

O *software* empregado para medidas empíricas está disponível nos arquivos desta dissertação, no endereço <<http://namakajiri.net/letras/mestrado2016>>, de forma livre e aberta ao público. Buscamos com isso estimular testes de reprodutibilidade, e também disponibilizar ferramentas auxiliares que foram criadas para nossa análise.

### 1.3.3 Coleta de exemplos

O terceiro método empregado para argumentação teórica foi o levantamento de contra-exemplos, no capítulo 5. Buscamos apresentar técnicas expressivas que são particulares ao sistema de escrita japonês, e que ilustram limitações do modelo transcricional da escrita (segundo o qual a escrita se limitaria a transcrever a linguagem falada). No estado atual do processamento automático de linguagens naturais, não é possível extrair tais exemplos automaticamente; por isso, a coleta foi manual e em pequena escala. Buscamos compensar essa limitação selecionando uma amostra eclética e representativa. Para demonstrar que tais técnicas são inerentes ao sistema, não limitadas a exceções ou malabarismos literários, extraímos exemplos de diversos períodos e registros linguísticos (literatura, cultura popular, textos acadêmicos, propaganda etc.). Esses exemplos foram organizados de forma a propor uma classificação e terminologia para técnicas de expressão da escrita em *kanji*.

## 1.4 Convenções

Por se tratar de um trabalho sintético sobre teoria linguística, empregamos grande número de convenções. Esta seção lista todas elas, como referência. Além disso, quando

<sup>19</sup> Agradeço ao prof. Jim Breen, do projeto EDICT, pela orientação na escolha de software.

uma notação técnica é empregada pela primeira vez, ela é acompanhada por uma nota (re)introduzindo-a brevemente.

Ao introduzir e definir um termo técnico, marcamos-lo em **negrito**. Os termos assim destacados estão reunidos no Glossário de Termos Técnicos (página 205).

Termos estrangeiros estão grafados em *itálico*. Exceto quando notado, os termos são do japonês. Quando do primeiro uso de cada expressão, uma tradução é oferecida entre aspas. Estas traduções estão reunidas no Glossário de Termos Estrangeiros (página 213).

Sons linguísticos estão representados pelo Alfabeto Fonético Internacional (IPA, sigla consagrada, do inglês *International Phonetic Alphabet*). Para uma descrição completa desta notação ver IPA (1999), ou manuais de introdução à linguística geral, como e.g. Tsujimura (2006). Neste padrão, transcrições fonêmicas são representadas entre barras, como em /gato/; e transcrições fonéticas, entre colchetes, como em [gato] (ver seção 3.1.1 para uma explicação do que são transcrições “fonéticas” e “fonêmicas”).

Letras impressas no papel normalmente representam sons ou palavras. Porém, em um trabalho sobre a escrita em si, com frequência é necessário discutir as próprias marcas gráficas, e não o que representam. Para fazer referência aos sinais gráficos, citamos-los entre *aspas angulares*, seguindo o padrão empregado por Rogers (ROGERS, 2004) e outros autores da área. Por exemplo: “a grafia de /konserto/ pode ser *⟨conserto⟩* ou *⟨concerto⟩*”. Porém, ao citar caracteres sino-japoneses em texto corrido em português, omitimos as aspas, por legibilidade. Por exemplo: “O caractere 生 tem 11 leituras diferentes.”

Para representar o japonês moderno no alfabeto romano, usamos o padrão estilo Hepburn (*Hebon-shiki*), a escolha comum em textos de língua europeia (HEPBURN, 1872). Os sons do japonês antigo estão descritos seguindo a reconstrução e a notação de Frellesvig (FRELLESVIG, 2010).

Embora popularmente se fale em “língua chinesa”, do ponto de vista linguístico o chinês não é uma língua, mas sim uma família de línguas (DEFRANCIS, 1984; MAIR, 1991). Nesta dissertação, o termo “chinês” é empregado neste sentido linguístico, como um termo genérico para a família inteira, em suas variações históricas e geográficas. Palavras do chinês em geral serão apresentadas em duas versões: na língua-padrão moderna (普通話 *Pǔtōnghuà*), baseada no mandarim (termo pelo qual a designaremos); e no chinês antigo (*Old Chinese*, 上古漢語 *Shàngǔ-hànyǔ*). O mandarim está representado na transcrição oficial da China, o *pīnyīn* (GB, 2012; SHIBLES, 1994), descrito no Apêndice C à página 215. O chinês antigo está transcrito seguindo a reconstrução e notação do *Old Minimal Chinese* de Schuessler (2007), escolhida por sua sobriedade (Schuessler limita-se às reconstruções mais bem estabelecidas e pouco controversas).

Traduções de morfemas (glosas morfológicas) seguem o padrão Leipzig (LEIPZIG, 2008; HASPELMATH; SIMS, 2002, 27–29). Um exemplo deste padrão se segue:

- (1.1) ゆく 河 の 流れ は  
*yuk-u kawa no nagar-e wa*  
 vai-ADN rio GEN fluir-NMLZ TOP  
 “O fluxo do rio que corre...”

O padrão Leipzig funciona da seguinte forma:

- A primeira linha apresenta a grafia japonesa. A escrita japonesa normalmente não usa espaços, mas na análise morfológica nós acrescentamos espaços separando as palavras. Usando os espaços, as palavras são alinhadas verticalmente nas três primeiras linhas.
- A segunda linha transcreve o japonês em romanização Hepburn (no caso do japonês clássico, transcrevemos a pronúncia moderna padrão). Os morfemas estão separados por hífen, e correspondem às glosas da linha 3.
- A terceira linha traduz morfema por morfema. A segmentação corresponde exatamente à da linha 2. Funções gramaticais são notadas pelas siglas do padrão Leipzig; as siglas usadas nesta dissertação estão reproduzidas à página 17. No exemplo acima, as siglas ADN, GEN, NMLZ e TOP denotam, respectivamente: “adnominal”, “genitivo”, “nominalizador”, e “tópico”.
- A quarta linha, entre aspas, apresenta uma tradução natural livre.
- Note que não separamos os morfemas por hífen na primeira linha. Isso é porque a escrita japonesa nem sempre permite separar os morfemas graficamente. Por exemplo, na frase acima, a palavra *yuk-u* tem uma raiz terminada em consoante, o que não é representável em escrita japonesa; os caracteres ゆく, *yu.ku*, correspondem cada um a uma sílaba inteira, de forma que é impossível separar *yuk-* na primeira linha. Também existem caracteres *kanji* que expressam mais de um morfema, como 雷 *kami-nari* divino-grito = “trovão”. Por isso, a primeira linha não é segmentada morfológicamente. Quando relevante, discutiremos no texto o que cada caractere representa.

## 2 Fundamentos teóricos: Duas teorias da escrita

### 2.1 A primazia da fala

Qual a relação entre a escrita e a fala? Este capítulo introduz duas abordagens teóricas opostas, que no decorrer desta dissertação serão contrastadas tomando o japonês como estudo de caso.

Segundo [Kuhn \(1996\)](#), o progresso da ciência se dá por uma sucessão de modelos de trabalho, chamados **paradigmas**.<sup>1</sup> Os paradigmas são estudos influentes, cujo método torna-se um exemplar a ser seguido por outros pesquisadores—um exemplo de como fazer ciência. Cada paradigma implica uma teoria que explica determinados fatos, e também um programa de pesquisa que norteará o que pesquisar daquele ponto em diante.

No paradigma atualmente dominante na linguística, associado a Chomsky e à escola gerativista, a linguagem é considerada uma capacidade inata da espécie humana, geneticamente pré-configurada na estrutura cerebral ([CHOMSKY, 2004](#); [PINKER, 2010](#)). Uma evidência importante para essa teoria é a aquisição linguística: todos os bebês adquirem espontaneamente qualquer língua a que sejam expostos, independente da habilidade individual, e sem precisar de nenhuma explicação explícita sobre a gramática ([GLEITMAN; NEWPORT, 1995](#)). Esse processo dito “natural” de aquisição, porém, só se dá através da fala sonora ou dos gestos manuais (no caso das línguas de sinais); e nunca através da escrita. A escrita precisa ser aprendida mais tarde, após a aquisição da fala, com instrução explícita e esforço consciente. Por isso, linguistas desta orientação rejeitam expressões como “linguagem escrita”; a escrita não faria parte da capacidade biológica inata chamada “linguagem”. A linguagem propriamente dita seria apenas a fala, definida como a capacidade natural que vemos manifesta em voz ou gesto; enquanto que a escrita seria meramente um artefato cultural de segunda ordem, uma representação indireta da fala ([ARONOFF, 1985](#), p. 28; [PINKER, 2010](#), p. 2).<sup>2</sup> Este princípio é chamado de **primazia da fala** ([LYONS, 1972](#); [DEFRANCIS, 1989](#), pp. 217–218).

O princípio da primazia da fala é sustentado pelo pressuposto biológico do paradigma gerativista. A proposta que a fala seria de alguma forma “primária” em relação à escrita possui ainda outros argumentos. [Lyons \(1972\)](#) reúne os principais:

<sup>1</sup> Nesta dissertação, o negrito denota termos técnicos, que podem ser consultados no Glossário próprio.

<sup>2</sup> Para uma opinião contrária dentro da linguística gerativa, ver [Aaron e Joshi \(2006\)](#).

- Todo ser humano precisa aprender a fala antes da escrita (primazia ontogenética);
- Toda cultura conhecida desenvolveu fala antes da escrita (primazia filogenética);
- A escrita seria uma representação da fala (primazia estrutural);
- E a fala serviria a propósitos mais amplos que a escrita (primazia funcional).

O modelo da primazia da fala, adotado pelo gerativismo, não se resume a essa escola. A ideia básica remonta a Aristóteles (século III A.C.), para quem “palavras faladas são símbolos da experiência mental, e palavras escritas são símbolos das palavras faladas” (ARISTÓTELES, 2009, parte 1). No estabelecimento da linguística moderna, no começo do século XX, Saussure definiu como objeto linguístico a palavra falada, e afirmou ainda que a única razão de ser da palavra escrita é representar aquela (SAUSSURE, 2008, p. 34). Este pressuposto se manteve entre os estruturalistas americanos; Bloomfield (1933) famosamente definiu que (p. 21)

A escrita não é linguagem, mas meramente uma forma de gravar a linguagem por meio de sinais visíveis.

Ou seja, a escrita seria, nas palavras do título do livro de DeFrancis sobre o tema, a fala tornada visível (DEFRANCIS, 1989). Hoje em dia, o modelo da primazia da fala permanece sendo o mais difundido na linguística em geral.

Podemos representar este modelo pelo seguinte diagrama:

**Figura 1 – Modelo da primazia da fala**



## 2.2 A autonomia da escrita

A postura teórica oposta à primazia da fala é a da **autonomia da escrita**, associada primariamente à escola funcionalista de Praga. Para os linguistas do paradigma funcional, a estrutura da linguagem está inter-relacionada às suas funções; e a função da escrita é radicalmente distinta da função da fala (VACHEK; LUELSdorFF, 1989; HALLIDAY,

2014). A fala, em seu uso natural, destina-se à interação conversacional e efêmera; enquanto que a escrita foi criada para a comunicação persistente, de um para muitos, com audiência indeterminada. Por exemplo, os sinais de parênteses podem ser entendidos como ferramentas para criar textos alternativos, em um esforço para atender às necessidades de diferentes audiências (Vachek, op.cit.; cf. também Nunberg (1990, pp. 104–). Considere as seguintes frases escritas:

- (2.1) “Os sinais de pontuação não necessariamente correspondem à prosódia (o ritmo e entonação da fala).”
- a. “Os sinais de pontuação não necessariamente correspondem à prosódia.”
  - b. “Os sinais de pontuação não necessariamente correspondem à prosódia, isto é, ao ritmo e entonação da fala.
- (2.2) “Os sinais de pontuação não necessariamente correspondem ao ritmo e entonação da fala (a prosódia).”
- a. “Os sinais de pontuação não necessariamente correspondem ao ritmo e entonação da fala.”
  - b. “Os sinais de pontuação não necessariamente correspondem ao ritmo e entonação da fala, isto é, à prosódia.”

No exemplo 2.1, se os parênteses forem deixados de lado (como em 2.1.a), temos um texto primário que assume que o leitor sabe o que é “prosódia”. No exemplo extenso, 2.1.b, o escritor acrescentou uma definição, e está portanto supondo que o leitor não conheça a palavra. O exemplo 2.1, com parênteses, é uma espécie de representação simultânea de a. e b. Para o leitor que conhece o sentido da palavra, os parênteses servem como um sinal de que aquela explicação é básica, que é seguro passar por ela sem dar muita atenção.

Já o exemplo 2.2 funciona de forma inversa. O autor assume que a audiência principal não conhece a palavra “prosódia”, ou que conhecê-la não é importante para aquele texto. Para esses leitores, a informação entre parênteses pode ser deixada de lado, criando uma leitura como 2.2.a. Mas, caso um leitor já conheça o conceito de “prosódia”, é útil apontar que é disto que o texto está tratando, e para esse leitor o texto equivale ao 2.2.b.

Em outras palavras, a escolha entre o que vai dentro e fora dos sinais de parênteses relaciona-se à audiência que o escritor tem em mente. Nesta e em outras técnicas, a escrita caracteriza-se por buscar a clareza imediata de sentido, a visão panorâmica (*surveyability*), a busca espacial de informação (Vachek, op.cit.). Para isso, a escrita escora-se no suporte bidimensional (o plano da página), fundamentalmente mais panorâmico

que a sequência linear de sons que perfaz a fala. Seria por esse motivo, por exemplo, que preferimos estudar linguística em livros-texto do que por meio de gravações de áudio, ainda que estas tenham se tornado tecnologicamente viáveis; o meio escrito é mais claro para se obter uma visão geral, para reler um parágrafo até compreendê-lo, para a busca de informações específicas. Esta distinção funcional faria com que, em sociedades letradas, a escrita evoluísse para além da simples representação da fala, tornando-se um sistema autônomo com características formais próprias (VACHEK; LUELSDORFF, 1989, p. 1–8; 56–60).

Também para M.A.K. Halliday, a escrita é um sistema linguístico de primeira ordem, não uma transcrição da fala. Em sua gramática funcional (HALLIDAY, 2014), Halliday divide a linguagem em “estratos”, dentre os quais estão a fonologia e a ortografia (“grafologia”)—dois elementos no mesmo nível, sem que a ortografia seja dependente da fonologia. Isso é porque, embora haja analogia entre a fala e a escrita, a analogia não é completa, e Halliday reconhece a existência de estruturas gramaticais próprias à escrita:

A escrita não é a representação do som da fala. Ainda que todo sistema de escrita esteja relacionado ao sistema sonoro de sua língua de forma sistemática, não-aleatória [...] essa relação não é direta. Há outro nível de organização na linguagem, e tanto o sistema sonoro quanto o escrito relacionam-se a ele. Trata-se do nível da lexicogramática.<sup>3</sup> O sistema sonoro e a escrita são os dois modos de expressão pelos quais a lexicogramática de uma língua pode ser representada, ou “realizada”.

Um vez que a linguagem evoluiu na espécie humana enquanto fala, todo sistema de escrita é originalmente parasítico à fala. [...] Mas, quando uma escrita evolui, ou, no nível individual, depois que uma criança em letramento termina por dominá-la e pode usá-la na prática, a escrita toma vida própria, alcançando diretamente a lexicogramática da linguagem ao invés de acessá-la através do som; e esse efeito é reforçado pela complementariedade funcional entre fala e escrita.

Na gramática funcional de Halliday, a “grafologia” sempre é considerada como uma alternativa em pé de igualdade com a “fonologia”, a tal ponto que a terminologia do autor redefine os termos “frase” [*clause*] como uma categoria abstrata da gramática, e “oração” [*sentence*] como uma estrutura específica da escrita, análoga à “frase” mas não idêntica (HALLIDAY, 2014, p. 6–8; 436).

Na mesma linha de Halliday, Geoffrey Nunberg descreve com mais detalhes algumas dessas estruturas abstratas próprias da escrita, analisando sua sintaxe específica, subjacente ao sistema de pontuação; e encontra ali diferenças importantes com a sintaxe da língua falada (NUNBERG, 1990). O autor observa que, mesmo que a escrita não seja parte da capacidade natural da linguagem, suas estruturas “artificiais” também podem

<sup>3</sup> Halliday propõe que, no nível profundo, não há distinção entre o léxico (o conjunto de palavras) e a gramática (as regras da língua); daí o termo.

ser analisadas com as ferramentas teóricas da linguística. Isso é possível porque, ainda que seja uma invenção cultural, a escrita é uma “aplicação” da capacidade biológica da linguagem (NUNBERG, 1990, p. 1–7):

Quero dizer com isto que o sistema [de estruturas específicas à escrita] surgiu como uma espécie de aplicação dos princípios de *design* da linguagem natural, tomados como ferramenta para impor alguma ordem no conteúdo lexical dos textos escritos. Quando visto como um todo, o sistema não tem análogos na língua falada (embora algumas de suas funções se assemelhem a vários mecanismos da fala). O sistema surgiu, no decorrer da padronização da escrita, em resposta às necessidades comunicativas específicas da escrita, e aproveitando-se dos recursos que apenas a notação gráfica oferece. A escrita é o que se segue, *grosso modo*, do ato de assentar a linguagem [*setting language down*].

Por exemplo, uma distinção simples que Nunberg ilustra como específica da escrita (adaptada ao português) é:

- (2.3) a. O chefe mandou um recado; todos os funcionários voltaram para casa.  
b. O chefe mandou um recado: todos os funcionários voltaram para casa.  
c. O chefe mandou um recado—todos os funcionários voltaram para casa.

No exemplo 2.3.a, a segunda frase, “todos os funcionários voltaram para casa”, nunca corresponde ao conteúdo do recado; no exemplo 2.3.b, ela precisa corresponder; e no exemplo 2.3.c, ambos os sentidos são possíveis. A distinção é marcada unicamente pelo sinal de pontuação, e não corresponde a nada na fala—ao contrário do que às vezes é sugerido, a pontuação não se limita a representar a prosódia falada. O argumento de Nunberg é que, embora esse tipo de notação não tenha correspondente na fala, ainda assim ele trabalha com estruturas sintáticas e semânticas tipicamente linguísticas, e que podem ser estudadas por técnicas linguísticas.

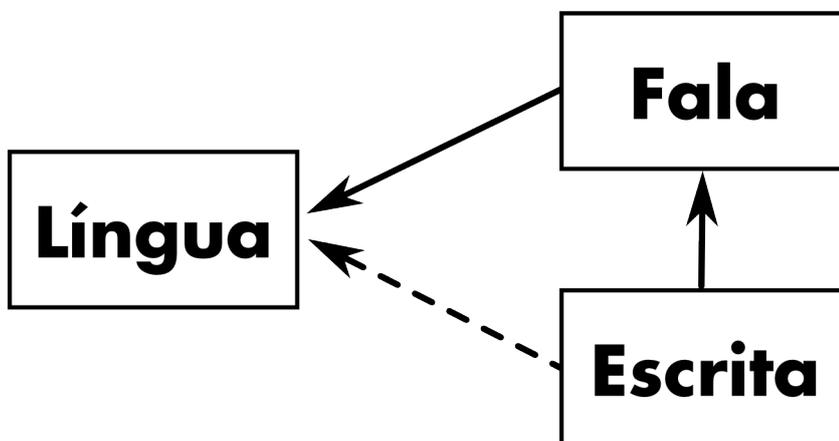
Nesta dissertação, classificaremos modelos como os de Vachek, Halliday e Nunberg como caracterizados pelo princípio da **autonomia da escrita**, em oposição à primazia da fala de Aristóteles, Bloomfield ou DeFrancis. Podemos comparar os dois modelos esquematicamente pelos diagramas:

Figura 2 – Comparação entre os modelos da primazia da fala e da autonomia da escrita

Modelo da primazia da fala:



Modelo da autonomia da escrita:



No diagrama do modelo da autonomia, as linhas sólidas representam o caminho original de significação: a escrita começa como a representação da fala, que por sua vez representa a língua. A linha tracejada representa o caminho que emerge em uma escrita madura: o meio escrito passa a ser capaz de referenciar sentidos linguísticos por mecanismos próprios, independentes da fala.

Os defensores da primazia da fala estão, é claro, cientes de que há elementos na escrita que não correspondem a nada na fala (tais como a pontuação, a tipografia, a distinção de homófonos, a organização bidimensional em tabelas, as notas de rodapé etc.). Similarmente, os teóricos da autonomia da escrita reconhecem a primazia ontogenética e filogenética da fala (ou seja, sua necessária antecedência na espécie e no indivíduo), e a origem da escrita como registro da fala. A diferença está na importância dada aos elementos que são específicos à escrita: no modelo da primazia da fala, tais elementos autônomos são considerados periféricos ou excepcionais, e a escrita seria fundamentalmente um registro em segunda mão da fala (primazia estrutural). Já para o modelo da escrita autônoma, a frequência e importância desses elementos provam que a escrita é uma realização da própria capacidade da *linguagem*—além da fala.

## 3 Som e sentido na escrita: a fonografia e a morfografia

Este capítulo tem o objetivo de descrever como funciona o sistema de escrita japonês, e qual a sua relação com a língua falada. Para isso, será proveitoso compará-lo às escritas alfabética e chinesa, que serão analisadas primeiro.

A escrita japonesa é baseada na combinação de dois tipos de sinais gráficos: os que representam apenas sons, e os que também especificam sentidos. Na seção 3.1, faremos uma recapitulação sobre as unidades linguísticas de som e sentido: os fonemas e os morfemas. Na seção 3.2 faremos uma descrição do alfabeto português, como exemplo de grafia pelo som. Na seção 3.3 estudaremos a escrita chinesa (fonte e base da japonesa) como uma introdução à grafia do sentido. Porém, como veremos, a grafia portuguesa também faz referência ao sentido, e a chinesa também se baseia no som; parece ser difícil ou impossível construir uma escrita usável que se escore puramente em um ou outro.<sup>1</sup> Por fim, na seção 3.5 apresentaremos o japonês como uma escrita “mista”—isto é, com especialização gráfica: um tipo de sinal visual depende apenas do som, e outro carrega também informação de sentido.

### 3.1 A dupla articulação da linguagem: fonemas e morfemas

Podemos considerar a linguagem como um sistema de **signos**, comparável ao sistema dos sinais de trânsito ou dos símbolos cartográficos. Um signo possui dois elementos: um **significante**, isto é, uma forma que o identifica; e um **significado** atribuído àquela forma. Por exemplo, nos mapas, o desenho  é um significante cujo significado é “aqui há um restaurante”. A palavra “restaurante”, no português falado, pode ser entendida como um signo semelhante, cujo significante (forma) é um padrão sonoro próprio a esta palavra, e cujo significado é a idéia ou conceito de restaurante. Este é o modelo **semiótico** introduzido por Saussure na aurora da linguística moderna (SAUSSURE et al., 2011).

Um sistema que permita a combinação de signos é dito **articulado**. Por exemplo, uma placa de trânsito com os sinais  combina o sentido de “restaurante” ao de “curva à direita”, gerando um sentido composto “restaurante virando à direita”. A linguagem, naturalmente, é articulada: podemos combinar as palavras em palavras

<sup>1</sup> Para uma discussão, ver Sampson (1994); DeFrancis e Unger (1994). Para um possível exemplo de escrita puramente simbólica, ver Rogers (2004), pp. 263–268, sobre o sistema artificial Blissymbols; e Unger (2004) para uma crítica.

maiores (como “parapeito”), frases, orações, períodos etc., que chamaremos genericamente de **enunciados**.

Além disso, se compararmos os signos da linguagem aos signos em geral, descobriremos nos signos linguísticos uma característica particular: seus significantes são compostos por componentes menores, recorrentes, que podem ser recombinaados infinitamente. Por exemplo, as palavras “ave”, “avô” e “avó” começam com os mesmos dois sons, sendo distinguidas apenas pelo terceiro. Note que, ao contrário dos significantes inteiros, esses sons individuais não carregam significado. Não é o caso que os sons “-e”, “-ô” e “-ó” nestas palavras signifiquem um pássaro ou um ancestral masculino ou feminino. A sequência sonora “avô” tem um significado, mas a unidade isolada “-ô” não. O fato de não carregarem sentido permite o reuso; por exemplo, os mesmos três sons que distinguem “ave”, “avô” e “avó” também distinguem “de”, “do” e “dó”.

Ou seja, a linguagem é articulada em dois níveis:

1. É composta por signos no sentido de Saussure, cada um com um significante (som) associado a um significado (sentido);
2. E, além disso, cada significante é composto por uma sequência de elementos (sons) menores, recorrentes, que não têm significado, mas que servem para distinguir os significantes uns dos outros.

Esta é uma característica econômica e produtiva: combinando um pequeno número de componentes unitários, podemos criar um enorme número de signos. Além disso, os próprios signos podem se recombinar em enunciados maiores, como frases, discursos e textos. A combinatória da linguagem é comparável talvez a um jogo de blocos de montar, no qual as peças representam unidades como “soldado”, “carrinho”, “árvore” etc.; e, além disso, essas peças em si são compostas por pecinhas menores, que não representam nada, mas podem ser reaproveitadas para criar novas peças. André Martinet famosamente chamou esta propriedade de **dupla articulação da linguagem** (MARTINET, 1984).

Os sons unitários que distinguem signos são chamados “fonemas”, e os signos elementares que carregam sentido são chamados “morfemas” (op.cit.). As escritas do mundo são baseadas na representação desses dois níveis de articulação. Para entender o que isto quer dizer, vamos rever, nas duas seções a seguir, como funcionam os fonemas e morfemas.<sup>2</sup>

<sup>2</sup> O que apresentamos aqui são definições tradicionais na linguística, que podem ser encontradas em qualquer manual introdutório; a nossa discussão é baseada em Fiorin (2008), IPA (1999), e Haspelmath e Sims (2002). Porém, estes modelos têm seus detratores, e algumas teorias recentes propõem modelos radicalmente diferentes para a fonologia e a morfologia. Para os propósitos de nosso trabalho, as definições tradicionais, que são as mais difundidas, nos serão suficientes.

### 3.1.1 Blocos elementares da linguagem: os fonemas

Considere as seguintes palavras do português:

- (3.1)
- Vá
  - Lá
  - Já
  - Chá

Ao pronunciar essas palavras, o som que produzimos no final de cada uma, o “a”, é idêntico. Descrito de outra forma, os gestos que produzimos para emitir essas palavras terminam num mesmo gesto básico, o de abrir bastante a boca e vibrar as cordas vocais. Já o som que precede o “a” é diferente em cada palavra, correspondendo a diferentes gestos. Por exemplo, na palavra “vá”, o lábio inferior apóia-se nos dentes enquanto o ar passa por eles; enquanto que na palavra “lá” é a ponta da língua que se movimenta, detendo-se no céu da boca atrás dos dentes frontais (a região do alvéolo). Essa diferença de gesto causa uma diferença no som, que serve para distinguir uma palavra da outra—ou seja, o som distinto muda o sentido de “vá” para “lá”.

Compare isso com o som final da palavra “mar”. Dependendo da região do Brasil, a pronúncia deste som é diferente. Na capital de São Paulo, a pronúncia mais prestigiada é com um toque da ponta da língua no alvéolo, como na palavra “Maria”. Mas, no interior do Estado, é mais comum um som produzido dobrando a ponta da língua para trás, sem tocar o céu da boca (popularmente conhecido como “R caipira”—embora este som também ocorra na capital). Já no Rio de Janeiro, a ponta da língua não é usada; ao invés disso, ergue-se o fundo da língua, alterando a passagem do sopro de ar, e gerando o “R carioca”.

Cada som distinto que uma língua usa é chamado um **fone** (do grego *phōnē*, “som”). Observe que, no exemplo da palavra “mar”, mudar os fones não mudou o sentido; embora os fones sejam bem diferentes, todos eles são entendidos como se referindo à mesma palavra. Ou seja, a palavra “mar” aceita todo um conjunto de sons (fones) distintos na última posição (o “R paulista”, “caipira”, “carioca”...). Um conjunto de fones que é desta maneira considerado como sendo “o mesmo significante” é chamado de **fonema**. Ou seja, um fonema é um conjunto de sons (fones) que têm o mesmo valor em uma determinada língua, como os vários sons de “R” do português. Observe que:

- No exemplo 3.1, “vá/lá/já” etc., trocar os sons mudou o sentido. Dizemos por isso que os fonemas foram alterados (ou seja, os fones pertencem a fonemas diferentes).

- No exemplo da palavra “mar”, trocar os sons não mudou o sentido. Dizemos que foram alterados os fones, mas sem alterar o fonema (ou seja, os fones pertencem ao mesmo fonema).

Os vários fones possíveis de um mesmo fonema são chamados **alofones** (do grego *állos*, “outro” + fone). A distinção entre um fone e outro é dita **fonética**, e entre um fonema e outro **fonêmica** (ou **fonológica**). Segue-se que distinções fonéticas não alteram o sentido, mas as fonêmicas, sim. No Alfabeto Fonético Internacional (IPA), o nível fonético é representado entre colchetes ([ ]), e o fonêmico, entre barras (/ /). Por exemplo:

- A diferença entre “vá” e “lá” é fonêmica, então aparece na notação entre barras: /va/ vs. /la/.
- A diferença entre as três pronúncias de “mar” é apenas fonética, então não aparece na notação entre barras; todas as variantes de pronúncia seriam representadas como /mar/. A diferença só aparece na notação em colchetes: “paulista”, “caipira” e “carioca” seriam [mar], [maɹ] e [max], respectivamente. Dizemos que [mar], [maɹ] e [max] são **realizações** possíveis de /mar/.

Cada língua pode empregar fones diferentes; por exemplo, a vogal portuguesa em “mão” não é usada em inglês, e a consoante inglesa de *think* não é usada em português. Uma variação mais sutil entre as línguas vem do fato que, mesmo quando os fones são os mesmos, cada língua pode traçar linhas distintas para agrupá-los em fonemas. Isto é, uma variação sonora que em uma língua é só fonética (que não muda o sentido) em outra pode ser fonêmica (mudar o sentido). Por exemplo:

- Em romeno, a palavra “dur” com um “R paulista”, [dur], significa “duro”; e a mesma palavra com um “R carioca”, [dux], significa “fantasma”. Em romeno, /r/ e /x/ (o “R carioca”) são fonemas diferentes; mas em português, [r] e [x] são alofones do mesmo fonema, /r/.
- Em japonês, o som “l” de “cala” pertence ao mesmo fonema que o “r” de “cara”, de forma que os sons [ka'ra] e [ka'la] são entendidos como variações da mesma palavra, *kara* (“a partir de; porque”). Ou seja, em português, /l/ e /r/ são fonemas diferentes, mas em japonês, [l] é apenas um alofone do fonema /r/.
- Em português, a palavra “tô” (forma coloquial de “estou”) pode ser pronunciada curta ou longa, sem mudar o sentido; ou seja, /to/ pode ser realizada indiferentemente como [to] ou [to:], onde [o:] simboliza um som mais longo. Em japonês, o som longo mudaria o sentido; por exemplo, de “porta” 戸 /to/ para “torre” 塔

/to:/. Em português, a duração das vogais não é fonêmica—não muda sentido—mas em japonês, sim.

### 3.1.1.1 Variação fonética condicionada e livre

Considere a palavra “fofo”. Seus fonemas são transcritos como /fofo/; mas, na pronúncia mais comum no Brasil, as vogais são realizadas de forma diferente no nível fonêmico, que pode ser transcrito como [fofɔ]. O último símbolo representa o som semelhante a um [u] que os brasileiros produzem<sup>3</sup> ao pronunciar o /o/ em fim de palavra. Isto é, o fonema /o/ possui dois alofones (duas realizações possíveis), [o] e [ɔ],<sup>4</sup> e podemos prever quando cada uma aparece:

1. Em final de palavra, /o/ se realiza como [ɔ]
2. Caso contrário, /o/ se realiza como [o]

Quando a variação fonética é determinada pelo contexto, como no exemplo acima, dizemo-la **condicionada**. Quando não há uma regra clara que permita prever qual será o alofone utilizado, a variação fonética é chamada **livre**. Descrevemos acima três variantes do /r/ em fim de sílaba, consideradas “carioca, paulistana e caipira” ([x], [r] e [ɾ]); mas não é raro que, por exemplo, um mesmo falante empregue tanto a variante “caipira” quanto a “paulistana”. Como não há um contexto linguístico simples que determine regularmente qual o alofone empregado, chamamos esse caso de variação livre. Muitas vezes a variação livre não é completamente “livre”, mas sim determinada por variáveis externas, como o formalismo da situação, a audiência, o dialeto etc.; mas isso parece ser um fenômeno diferente do condicionamento linguístico simples, como o caso do /o/ em fim de palavra.

### 3.1.2 Elementos básicos de sentido: os morfemas

No começo da seção 3.1 (p. 47), descrevemos a linguagem (seguindo [Saussure](#)) como um sistema de signos, isto é, pares de significante e significado. Até agora, temos usado as palavras como exemplos de signos. Contudo, as palavras podem ser signos compostos; elas não são os signos mais fundamentais da linguagem. Considere:

- (3.2)
- Paraquedas
  - Para-brisa
  - Para-lama
  - Para-raios

<sup>3</sup> Exceto em alguns dialetos, como os do Sul.

<sup>4</sup> Por que um símbolo [ɔ] distinto de [u]? Quem é leitor do português pensa nesse som em termos da letra “u”; mas, se ouvimo-lo com atenção, notaremos que ele não é tão pronunciado quanto a vogal final de e.g. “caruru”. O símbolo [ɔ] representa uma vogal cujo som é intermediário entre [o] e [u].

Claramente essas palavras são compostas por mais de uma unidade de sentido. Todas começam incorporando o signo “para”; ou seja, todas começam pelos fonemas (significante) /para/, que em todas elas carrega o significado de “parar, interromper, proteger de”.

Podemos entender tais palavras como sendo composições de duas outras. Na ortografia portuguesa, essa composição pode ser indicada pelo sinal de hífen, ou não, de forma arbitrária; dentre as palavras no exemplo 3.2 acima, apenas “paraquedas” não leva hífen, segundo edição atual das normas ortográficas (ABL, 2009)—embora claramente a estrutura de “paraquedas” seja perfeitamente análoga às demais.

Agora considere as seguintes palavras:

- (3.3) • Paraquedismo  
• Trapezismo  
• Maratonismo

- (3.4) • Paraquedista  
• Trapezista  
• Maratonista

Os exemplos em 3.3 referem-se todos a alguma atividade ou prática, e todos terminam em “-ismo”. Já os exemplos em 3.4 referem-se todos aos respectivos praticantes ou aderentes, e as palavras são semelhantes, sempre trocando “-ismo” por “-ista”.

Ao contrário de “para” ou “quedas”, “-ismo” e “-ista” não são palavras: estas sequências não aparecem sozinhas, mas somente como parte de palavras maiores. Mas, claramente elas são signos: são sequências de fonemas iguais que carregam um mesmo sentido. Ou seja, podemos dividir a palavra “paraquedismo” em dois signos, e novamente subdividir o primeiro signo:

**Tabela 3 – Subdivisão da palavra “paraquedismo” em signos**

Significante	Significado
paraqued-	“paraquedas”
para-	“parar algo, proteger de”
qued-	“queda”
-ismo	“prática de”

Note que cada um desses signos aparece em outras palavras, carregando o mesmo sentido: “qued-”, por exemplo, aparece em “queda”, “quedar”, “quedando” etc. Porém, não é possível dividir “qued-” em signos menores. Não é o caso que “que-” ou “d-” carreguem um sentido próprio, recorrente em outras palavras.

Ou seja, sequências de fonemas como “para” e “qued-” são sequências mínimas que carregam sentido. Qualquer sequência menor não têm sentido próprio, i.e. não é um signo.<sup>5</sup> Chamamos tais signos mínimos de **morfemas**. Um morfema pode ser descrito como uma sequência de fonemas tal que:

1. A sequência possui sentido próprio, e
2. Nenhuma sub-sequência possui sentido próprio.

Morfemas como “para”, que ocorrem como palavras independentes, são ditos **livres**. Morfemas como “-ista”, que só aparecem como parte de outras palavras, são ditos **presos**.<sup>6</sup>

Quando queremos chamar a atenção para os morfemas de uma palavra, separamos por hífens, independentemente da norma ortográfica (cf. padrão Leipzig à página 40). Por exemplo, “para-qued-ista”.

Algumas palavras possuem variações em seu som e sentido, que são presentes também em outras palavras da mesma categoria. Considere:

(3.5) • Gato

• Gata

• Gatos

• Gatas

(3.6) • Menino

• Menina

• Meninos

• Meninas

(3.7) • Afinado

• Afinada

• Afinados

• Afinadas

---

<sup>5</sup> Ver contudo nota sobre simbolismo sonoro à página 33.

<sup>6</sup> Todos os termos e conceitos recapitulados nesta seção podem ser encontrados com explicações mais detalhadas na op. cit., [Haspelmath e Sims \(2002\)](#), ou em outros manuais de morfologia.

- (3.8)
- Desafinado
  - Desafinada
  - Desafinados
  - Desafinadas

Todas as palavras acima possuem uma sequência idêntica no começo, como “gat-” ou “menin-”, e uma parte final que varia, ou **flexiona-se**, segundo um mesmo esquema, ou **paradigma**: “-o, -a, -os, -as”. A parte de uma palavra que não varia é chamada **base** ou **radical**. A base pode ser composta por um só morfema, como “gat-” ou “menin-”, ou por mais de um, como “des-a-fin-” (compare “fino”). Uma base elementar, de um só morfema, é chamada **raiz**. Dizemos, por exemplo, que a base (ou radical) “des-afinad-” é composta a partir da raiz “fin-”.<sup>7</sup> Os morfemas presos que se acoplam às raízes são chamados **afixos**; um **sufixo** é um afixo que aparece depois da base, como “-a” e “-s” em “gata/gatas”; enquanto um **prefixo**, como “des-”, é um afixo que aparece antes.

Em alguns casos, é útil considerar sequências fonológicas distintas como sendo variações do mesmo morfema. Considere o pronome átono oblíquo “a”, do português clássico,<sup>8</sup> quando se segue ao verbo:

(3.9) Amar + a:

- Amei: amei-a
- Amo: amo-a
- Amava: amava-a
- Amar: amá-la
- Amaria: amá-la-ia
- Amaram: amaram-na

<sup>7</sup> Fora do contexto da linguística, os termos “raiz” e “radical” são às vezes usados de forma pouco precisa ou intercambiável. Nesta dissertação, preferiremos o termo “base” a “radical”, pois o segundo confunde-se facilmente com “raiz”. Em inglês os termos para “radical”, “base” e “raiz” são *stem*, *base* e *root*, respectivamente; em japonês a base é chamada 語幹 *gokan* “tronco/caule da palavra”, e a raiz, 語根 *gokon* “raiz da palavra”.

<sup>8</sup> Em oposição ao português brasileiro moderno, que prefere o pronome “ela” em todos os casos.

## (3.10) Merecer + a:

- Mereci: mereci-a
- Mereço: mereço-a
- Merecia: merecia-a
- Merecer: merecê-la
- Mereceria: merecê-la-ia
- Mereceram: mereceram-na

Observe que normalmente o pronome átono feminino manifesta-se como ‘a’. Porém, quando a base flexionada do verbo termina em /m/, surge um som /n/ entre verbo e pronome. E quando a base termina em /r/, há toda uma transformação: o /r/ desaparece, a vogal da base que o precedia recebe acento, e surge um /l/ antes do ‘a’. Parece pouco razoável descrever /a/, /la/ e /na/ como morfemas distintos; é mais simples pensar nas três formas como variantes do pronome ‘a’, selecionadas por contextos fonológicos distintos. Tais variações fonológicas de um fonema são chamadas **alomorfes**—“outras formas”, em analogia aos “outros fones” (alofones) dos fonemas (seção 3.1.1 à página 49). Os processos que alteram a pronúncia de um morfema (ou seja, que selecionam alomorfes) são ditos **morfofonêmicos**.

Um exemplo deste fenômeno no japonês acontece com o morfema *ame* “chuva”, que, como exemplificado na seção 1.1 (p. 27), possui três alomorfes: *ame*, *ama-* (como prefixo, e.g. *ama-yadori* “abrigo da chuva”) e *-same* (como sufixo, e.g. *ko-same* “pequena chuva”), todos representados pelo mesmo caractere 雨.<sup>9</sup>

## 3.1.3 A dupla articulação e a escrita

Vimos que a linguagem pode ser descrita como um sistema de signos que se articulam em dois níveis, os fonemas e os morfemas. A escrita é um mecanismo para representar a linguagem por meio de sinais gráficos (SPROAT, 2000; ROGERS, 2004). Existem dois mecanismos básicos para isto, baseados nos dois níveis de articulação da linguagem:

- Representar, por meio de sinais gráficos, as unidades elementares desprovidas de sentido—isto é, os fonemas;

<sup>9</sup> Esta alomorfia é um resquício de um sistema morfofonêmico do japonês antigo, que causava mudanças vocálicas ou *ablaut* em posição presa; cf. *fune* “barco” / *funa-bito* “pessoa do barco, passageiro”; *sake* “bebida alcoólica” vs. *saka-ba* “lugar de bebida, adega”; *te* “mão” vs. *ta-moto* “[na] base da mão = manga”, etc. (MILLER, 1967, pp. 185–187). Esse sistema está fossilizado em certas palavras compostas, mas ele não é mais produtivo—ou seja, este tipo de alomorfia não acontece mais em palavras novas: e.g. *sake-kasu* “borra de saquê”, *te-gaki* “manuscrito”, etc.

- E representar integralmente morfemas; isto é, signos, tanto o significante (sequência de fonemas) quanto o sentido.

No restante deste capítulo, ilustraremos esta afirmação por meio de estudos de caso. Na seção 3.2, analisaremos a escrita do português como um exemplo de sistema baseado em fonemas. Na seção 3.3, introduziremo a escrita chinesa como um exemplo de representação baseada nos morfemas. E por fim, na seção 3.5 apresentaremos a escrita japonesa como mista—isto é, como uma escrita que possui sinais gráficos especializados para cada um dos dois níveis.

Por que analisar as escritas portuguesa e chinesa? Ocorre que os fenômenos, teorias e conceitos que se aplicam a essas escritas também são encontrados na japonesa. É útil ver primeiro como ocorrem nestas escritas mais simples, para então estudar como interagem no sistema japonês.

## 3.2 Uma escrita fonográfica: o português

Considere as seguintes palavras do português, e compare sua ortografia (transcrita abaixo entre aspas angulares, ⟨ ⟩) com os sons do português falado (transcritos no Alfabeto Fonético Internacional, IPA):

**Tabela 4 – Comparação da escrita portuguesa e variação fonológica**

/gato/

⟨gato⟩

/mato/

⟨mato⟩

/bato/

⟨bato⟩

/lato/

⟨lato⟩

Em cada exemplo, a palavra difere das demais apenas no som inicial; e essa diferença é fonêmica, ou seja, muda o sentido (cf. seção 3.1.1 à página 49). Agora considere a ortografia. Está claro que a cada fonema corresponde uma letra, e que, quando o fonema muda, muda também a letra correspondente. Este é o princípio básico do **alfabeto latino** (ou **romano**), empregado no português e em muitas outras línguas: cada letra representa um fonema. (Incidentalmente, é por isso que a notação IPA é tão visualmente semelhante ao português: ela também foi baseada no alfabeto latino. Mas não estamos falando aqui da coincidência dos sinais gráficos. Considere os sinais entre barras como uma forma de se referir ao som da fala, como se fosse um arquivo de áudio

em um documento eletrônico; o importante para nossa discussão é notar como cada letra representa um som—e um gesto que produz esse som.)

Agora observe as variações fonéticas (não fonológicas) da palavra “mar”, discutidas na seção 3.1.1, e reproduzidas abaixo junto à ortografia:

**Tabela 5 – Comparação da escrita portuguesa e pronúncia fonética**

[mar] (“R paulista”)  
⟨mar⟩

[max] (“R carioca”)  
⟨mar⟩

[maɹ] (“R caipira”)  
⟨mar⟩

Note que não há vestígio na escrita da variação em nível fonético. Quando introduzimos a variação fonética na seção 3.1.1, tivemos que usar circunlóquios como “o som popularmente conhecido como R caipira, com a ponta da língua dobrada para trás”, etc., porque não há maneira de representar especificamente este som na escrita portuguesa. Isso acontece porque, como argumentam [Vachek e Luelsdorff \(1989\)](#) e outros, a escrita está interessada sobretudo em transmitir o sentido de maneira clara; então, ela atenta principalmente ao nível fonológico (que interessa para o sentido) e não aos detalhes fonéticos (que são irrelevantes para a clareza semântica).

Chamamos de **princípio fonográfico** o método de grafar uma língua representando seus fonemas. A representação segundo este princípio é chamada **fonografia**, e uma escrita baseada (principalmente) na fonografia, como o português, é classificada como **escrita fonográfica** ([ROGERS, 2004](#)).<sup>10</sup> Um símbolo que representa fonemas é um **fonograma**; os fonogramas são um de vários tipos possíveis de caracteres ou sinais da escrita, que são chamados “grafemas”, como veremos na seção a seguir.

Algumas escritas fonográficas têm por base representar apenas alguns fonemas; por exemplo, o princípio do árabe é a representação das consoantes. Em outras, cada marca gráfica corresponde a vários fonemas; os símbolos do Cherokee representam uma sílaba inteira, enquanto os *kana* japoneses representam uma sílaba curta (“mora”; cf. sec. 3.5.5 à página 115). A escrita latina, usada no português e muitas outras línguas, tem por base o princípio que a cada fonema corresponde um símbolo gráfico distinto, e vice-versa. Esta regra se denomina o **princípio alfabético**, e um sistema de escrita nele baseado é um **alfabeto**, como o latino ou o grego (de onde veio o nome, a partir das duas primeiras letras:  $\alpha$  *alpha* e  $\beta$  *bêta*).

<sup>10</sup> Para uma maneira mais sofisticada de classificar sistemas de escrita, cf. [Sproat \(2000\)](#).

### 3.2.1 Grafos e grafemas

Em todas as palavras exemplificadas na tabela 4 acima, o fonema /a / é representado pela letra ⟨a⟩. Observe, porém, as seguintes formas gráficas:

Figura 3 – Variantes da letra ⟨a⟩



A primeira forma é chamada “A maiúsculo”, “capitular” ou de “caixa alta”; a segunda é um “A minúsculo romano serifado”, e a terceira, um “A itálico”. Observe que o desenho de cada letra é bastante diferente. Porém, por convenção, todas essas formas são processadas durante a leitura como variantes do mesmo signo, a “letra A”.

No estudo dos sistemas de escrita, chamamos cada marca visualmente distinta de **grafo**, e o significante abstrato que elas denotam de **grafema**, em analogia à distinção fone/fonema (cf. seção 3.1.1, p. 49).<sup>11</sup> Ou seja, na figura 3 acima, as três formas são grafos alternativos do grafema ⟨A⟩—ou **alógrafos** do grafema ⟨A⟩ (ROGERS, 2004, sec. 2.1.2).

Assim como no caso dos fones e fonemas, o que é ou não alógrafo é uma convenção, e depende das regras de cada sistema de escrita. Vimos acima três marcas muito diferentes, mas que são todas “a mesma letra” (assim como os três alofones de “mar” da seção 3.1.1 são considerados “a mesma palavra”). Pode-se argumentar que letra ⟨o⟩ é mais semelhante visualmente ao “A itálico”, ⟨a⟩, do que este com o “A maiúsculo”; mas o que está em jogo não é a semelhança gráfica, é o sistema de convenções, e no alfabeto latino moderno, ⟨o⟩ é um grafema distinto de ⟨a⟩, mas ⟨A⟩ não. Já no Alfabeto Fonético Internacional (IPA), diferentemente do português, as formas ⟨a⟩ e ⟨a⟩ são grafemas distintos; na tabela do IPA, o símbolo /a / representa o fone “vogal aberta anterior” (o som “A” do português, produzido com a língua mais para a frente), enquanto que o símbolo /a / representa a “vogal aberta posterior” (o som “A” do inglês britânico, com a língua mais para o fundo).

Nesta dissertação, quando não for necessário trabalhar com a distinção grafo / grafema, usaremos os termos comuns “letra”, para se referir aos grafemas do alfabeto, e “caractere”, para se referir aos grafemas chineses e japoneses.

<sup>11</sup> Na ciência da computação, o termo **glifo** (*glyph*) é usado com o mesmo sentido que o “grafo” da linguística (UNICODE, 2016).

### 3.2.2 Além do princípio fonográfico

Argumentamos até agora que a escrita portuguesa é basicamente fonográfica; ou seja, seu mecanismo fundamental é a representação de fonemas por meio de grafemas. Além disso, vimos que a escrita portuguesa é um alfabeto: normalmente cada grafema corresponde a exatamente um fonema, e vice-versa.

Contudo, quando falamos em “princípio”, queremos dizer a regra básica do sistema; não uma regra pétrea, que nunca varia. De fato, as escritas do mundo parecem todas seguir vários princípios em diferentes graus. Veremos nesta seção casos em que a escrita portuguesa não segue completamente o princípio alfabético, ou mesmo o princípio fonográfico.

#### 3.2.2.1 Dígrafos e plurigrafia

Considere as seguintes palavras, e compare a ortografia à pronúncia:

- (3.11) ⟨jato⟩ /ʒato /  
 ⟨chato⟩ /ʃato /

Quanto ao som, ambas as palavras possuem o mesmo número de fonemas. Na grafia, porém, a segunda usa mais grafemas. Isso ocorre porque o som /ʃ/ está representado por duas letras, ⟨ch⟩. Uma combinação de duas letras que representa um só som é chamada **dígrafo**. Em alguns sistemas de escrita, mais de dois grafemas podem ser usados; por exemplo, em francês, a sequência ⟨eau⟩ representa o som /o/. Generalizando, denominamos **plurigrafia** o uso de vários grafemas para representar algo que, pelo princípio do sistema, normalmente seria representado por um só.<sup>12</sup>

#### 3.2.2.2 Plurifonia

Considere agora a seguinte palavra, comparando ortografia e pronúncia, e observe o destaque em negrito:

- (3.12) ⟨nexo⟩ /nɛkso /

Este é o contrário do caso anterior; aqui, a sequência de dois fonemas /ks/ é representada por um só grafema, ⟨x⟩. Chamaremos casos como este de **plurifonia**.

<sup>12</sup> Esta definição é semelhante à de Rogers (2004), p. 16, mas mais generalizada. Evitamos o termo de Rogers, “poligrafia”, por conflitar com o uso que Boltz (1994) e outros fazem deste termo, que equivale ao que chamamos “polivalência” na seção 3.2.2.3 abaixo. No uso de Rogers, “poligrafia” faz analogia ao termo comum “dígrafia”; e no de Boltz, ao termo “polissemia”. Como não parece haver consenso terminológico na área, optamos por evitar termos com várias definições conflitantes. Agradeço a Wolfgang Lipp pela sugestão terminológica.

Tanto a plurifonia quanto a plurigrafia são casos de pluralidade no mapeamento entre escrita e língua. Podemos compará-las da seguinte forma:

**Tabela 6 – Pluralidade na escrita: plurigrafia e plurifonia**

	<b>Plurigrafia</b>	<b>Plurifonia</b>
<b>Sentido</b>	Escrita (da língua para o texto)	Leitura (do texto para a língua)
<b>Mapeamento</b>	1 fonema : N grafemas	1 grafema : N fonemas
<b>Exemplo</b>	/ʃa/ → ⟨chá⟩	⟨nexo⟩ → /nɛkso/

### 3.2.2.3 Polivalência

Considere as seguintes palavras:

(3.13) ⟨se⟩ /se/  
 ⟨use⟩ /uze/

(3.14) ⟨ri⟩ /ri/  
 ⟨guri⟩ /guri/

(3.15) ⟨xerox⟩ /ʃerɔks/  
 ⟨exame⟩ /ezame/

Observe, no destaque em negrito, que as mesmas letras podem representar sons diferentes. O ⟨s⟩ representa um som de /s/ no começo de palavra, mas um som de /z/ quando entre vogais. O ⟨r⟩ representa o “R forte” em posição inicial de palavra, e o “R fraco”, [r], no meio.<sup>13</sup>

Trata-se portanto de uma complicação no sistema, uma irregularidade do ponto de vista do princípio fonográfico. Normalmente, cada letra ou grafema representa sempre o mesmo fonema; mas estas letras têm vários valores possíveis. Chamamos este fenômeno de **polivalência**.<sup>14</sup>

A polivalência é um fenômeno diferente da pluralidade que vimos nas seções 3.2.2.1 e 3.2.2.2. No caso da leitura, por exemplo, um grafema plural representa vários fonemas em sequência—como o ⟨x⟩ quando representava /ks/. Diferentemente disto, um grafema polivalente possui vários valores em potencial; mas, em cada ocorrência, apenas um deles é selecionado. A letra ⟨s⟩ pode representar /s/ ou /z/, mas cada uso do ⟨s⟩ representa ou um ou outro.<sup>15</sup>

<sup>13</sup> A realização fonética exata do “R forte” varia bastante com o dialeto, mas sempre há uma distinção fonológica entre o “R forte” e o “fraco” (e.g. entre “carro” e “caro”).

<sup>14</sup> M.O. Connor usa a palavra neste sentido em [Bright e Daniels \(1996\)](#) (p. 94). Boltz, na mesma obra, usa o termo equivalente “multivalência” (*multivalence*).

<sup>15</sup> Para aqueles familiares com os conceitos de Jakobson, a pluralidade pode ser definida como um mapeamento um-para-muitos no eixo sintagmático, enquanto que a polivalência é uma multiplicidade no eixo paradigmático.

Casos de polivalência fonológica muitas vezes são consequências de mudanças na sonoridade nas línguas. A língua escrita parece ser conservadora por natureza; uma vez que os leitores adquirem um sistema de representação ortográfico, eles resistem a mudanças grafêmicas.<sup>16</sup> A fonologia, em contraste, muda constantemente, segundo fatores bem estudados (ver e.g. Labov (1965) e obra subsequente do autor). Para se apropriar da frase de Caius Titus, *verba volant, scripta manent* (“a fala voa, a escrita se mantém”). Assim, por exemplo, embora o som /h/ em começo de palavra tenha caído do latim bem cedo na história da língua, a letra ⟨h⟩ em começo de palavra se manteve até hoje (HERMAN, 2000; WRIGHT, 1982).

Na história das línguas, muitas mudanças fonológicas são **condicionadas**, no sentido visto na seção 3.1.1.1 (p. 51); isto é, são determinadas pelo contexto linguístico. Como reflexo disso, surge na escrita a **polivalência condicionada**. Por exemplo, na evolução do latim falado para o português, o que era inicialmente um som /s/, grafado ⟨s⟩, transformou-se no som /z/ quando entre vogais; porém, devido à resistência da escrita à atualização, o novo som /z/ continuou sendo grafado ⟨s⟩, e assim se mantém na ortografia portuguesa até hoje. Desta forma, uma aparente irregularidade da ortografia revela-se uma reprodução da história da língua; a regra “a letra ⟨s⟩ entre vogais soa como /z/” é uma descrição tanto da ortografia quanto da evolução das línguas românicas (HERMAN, op.cit.).

Nem toda polivalência da escrita é condicionada:

- (3.16) ⟨lixa⟩ /lif<sub>a</sub> /  
 ⟨fixa⟩ /fiks<sub>a</sub> /  
 ⟨exame⟩ /ezame /  
 ⟨próximo⟩ /prɔsimo /

A letra ⟨x⟩ do português é bastante polivalente na leitura. Em alguns casos essa polivalência tem algum condicionamento; por exemplo, um ⟨x⟩ em começo de palavra sempre é pronunciado /ʃ/, e em final sempre /ks/, donde ⟨Xerox⟩ tem uma pronúncia condicionada. Mas casos como ⟨lixa⟩ vs. ⟨fixa⟩ mostram que muitas vezes a escolha da pronúncia é **não-condicionada**; ela precisa ser memorizada palavra a palavra.

O que vimos até agora foi polivalência de leitura: um grafema apresentando múltiplas leituras em potencial. A polivalência também acontece no sentido da escrita:

<sup>16</sup> Certa vez, em uma aula de português para estudantes de nono anos, o autor discutia a ortografia polivalente da letra ⟨c⟩, e levantou como exemplo o fato que uma representação mais lógica (isto é, fonograficamente mais direta) da palavra “conhecimento” seria ⟨konhesimento⟩. A sala reagiu com reclamações de que “mas assim fica muito feio!”. Podemos especular que essa reação bastante universal de “feio” aconteça quando a representação gráfica não corresponde àquela já condicionada no cérebro; e que a mesma sensação age como um freio que retarda a atualização da ortografia—por vezes em níveis extremos, como no inglês e no francês, cujas escritas correspondem às fonologias de séculos atrás.

um fonema (ou sequência) pode ter várias representações alternativas. Em português, por exemplo, o som /s/ pode ser grafado de muitas formas:

- (3.17) /sem/ <sem>; <cem>  
 /asem/ <assem>  
 /aso/ <asso>; <aço>  
 /tras/ <traz>; <trás>  
 /dese/ <desce>  
 /esita/ <excita>

Novamente, há casos em que a polivalência é condicionada, em geral refletindo mudanças históricas: e.g. a letra <z> só é uma possibilidade para representar o som /s/ em final de sílaba; o grafo <s> singular nunca representa /s/ entre vogais, devido ao vozamento que vimos acima; o <c> só representa /s/ antes das vogais frontais (/i/, /e/, /ε/), refletindo a palatalização do <c> = antigo /k/ latino, etc. E também há casos de polivalência não-condicionada, como <sem>/<cem>, <asso>/<aço>, etc. (com frequência devido a mudanças históricas que resultaram em fusão de sons outrora distintos).

Resumindo o que vimos até agora, tanto a polivalência (multiplicidade em potencial) quanto a pluraridade (multiplicidade em sequência) podem ocorrer no sentido da leitura ou da escrita, e podem ser condicionadas ou não. A tabela abaixo esquematiza as possibilidades:

**Tabela 7 – Multiplicidade na escrita: pluraridade e polivalência**

Tipo de multiplicidade	Sentido	Nome	Exemplo
Em sequência	Leitura	Plurifonia	<x> → /ks/ em <nexo> → /nɛksɔ/
	Escrita	Plurigrafia	/ʃ/ → <ch> em /ʃa/ → <chá>
Em potencial	Leitura	Polivalência de leitura	A letra <x> pode ser lida: /ʃ/, /s/, /z/, /ks/
	Escrita	Polivalência de escrita	O som /s/ pode ser escrito: <s>, <ss>, <c>, <ç>, <xc>, <sc>

## 3.2.2.3.1 Polivalência e profundidade ortográfica

Considere a pronúncia das seguintes palavras:

**Tabela 8 – Morfofonologia do morfema *medic-***

medicamento	/me.di.ka'mẽ.to/
medicina	/me.di.'si.na/
medicinal	/me.di.'si.naw/
medicou	/me.di.'kow/
médico	/'mɛ.di.ko/

Na discussão até agora, vimos que a letra ⟨c⟩ é polivalente, possuindo duas pronúncias condicionadas: /k/ antes de /a/, /o/, /ɔ/, /u/; e /s/ antes de /i/, /e/, /ɛ/, refletindo desenvolvimentos históricos do latim (especificamente, o processo de palatalização). Mas deixemos de lado por um momento tanto a escrita quanto a história, e pensemos do ponto de vista sincrônico. Considere as palavras acima do ponto de vista de uma criança ou estrangeiro, que está aprendendo a língua sem saber nada ainda da escrita ou da etimologia. Observe que, nos sons da fala, o morfema *medic-* varia a pronúncia dependendo do sufixo, terminando ora com /k/, ora com /s/, de acordo com vogal seguinte. Este fenômeno independe da ortografia; por exemplo, mesmo para um cego congênito que não tenha sido ensinado sobre a letra ⟨c⟩, continua sendo verdade que o *som* nesta posição varia entre /k/, antes de /a,o,u/, e /s/, antes de /e,i/. Como vimos na seção 3.1.2 (p. 51), variações no som de um morfema são chamadas “alomorfes”, e o processo que muda a pronúncia de um morfema é dito “morfofonêmico”. Ou seja, o morfema *medic-* possui dois alomorfes, /medik/ e /medis/, condicionados pela vogal seguinte.

Tanto a polivalência da letra ⟨c⟩ quanto a alomorfia do morfema *medic-* têm origem no mesmo evento histórico: a palatalização do /k/ latino antes de vogais frontais. Porém, os dois tipos de polivalência são fenômenos distintos. Considere que não há relação morfofonêmica entre “saca” e “saci”; o uso da letra ⟨c⟩ nos dois casos é coincidência, e limita-se às regras da ortografia. Mas, no caso de *medic-*, o processo que alterou a pronúncia da letra ⟨c⟩ também alterou simultaneamente a pronúncia do morfema; de forma que *medic-* terminou por ter uma mesma forma gráfica para seus dois alomorfes. É possível argumentar que, neste caso, que a letra ⟨c⟩ não está representando a língua somente no nível fonêmico, mas que se tornou uma representação mais abstrata, atingindo o nível morfofonêmico. O mesmo ocorre com outros morfemas latinos que terminam em /k/ (fármaco/farmacêutico/farmácia; difícil/dificuldade; fonética/foneticista; autêntica/autêntico/autenticidade...)

Quanto uma escrita basicamente fonográfica alcança o nível morfofonêmico, dizemos que ela possui **profundidade ortográfica** (ROGERS, 2004, sec. 9.7). A profundi-

dade ortográfica é bastante comum nas escritas fonográficas; ou seja, a maioria delas não transcreve tão-somente fonemas, mas faz alguma referência a níveis mais abstratos da linguagem, unificando a representação de alomorfes. Como o próprio princípio fonográfico, isso é consequência natural do fato das escritas objetivarem principalmente a clareza de sentido. O grau exato de profundidade varia com cada sistema. Comparando aqueles que empregam o alfabeto latino, por exemplo, o finlandês é mais raso que o português, atendo-se bastante ao princípio fonográfico. Já o inglês é mais profundo: uma grande quantidade de morfemas têm alomorfes distintos grafados de forma idêntica, mesmo que isso viole o princípio fonográfico, como em ⟨sane⟩ /sem/ vs. ⟨sanity⟩ /'sænəti/, ou ⟨serene⟩ /sə'ri:n/ vs. ⟨serenity⟩ /sə'renəti/.<sup>17</sup>

O ponto importante desta discussão é que uma polivalência que, à primeira vista, parece ser uma irregularidade arbitrária na fonografia, como a letra ⟨c⟩ no português, pode acabar se revelando uma representação regular de um fenômeno que existe na língua falada: os processos morfofonêmicos. Em tais casos de ortografia profunda, uma escrita normalmente fonográfica está, até certo ponto, representando também morfemas; e aproximando-se portanto das escritas morfográficas, como o chinês e o *kanji* japonês (cf. seções 3.3, 3.5 a seguir).

#### 3.2.2.3.2 Polivalência e sentido: heterografia

Considere as seguintes palavras ortográficas:

(3.18) ⟨sem⟩ / ⟨cem⟩

(3.19) ⟨ouve⟩ / ⟨houve⟩

(3.20) ⟨viagem⟩ / ⟨viajem⟩

No exemplo 3.18, o som /s/ é representado ora pela letra ⟨s⟩, ora pela ⟨c⟩. Trata-se portanto de polivalência da escrita. Porém, observamos que essa polivalência foi cooptada para distinguir na escrita palavras que, na fala, partilham um significante idêntico: /sem/ “preposição de ausência” vs. /sem/ “cento”.

Como vimos anteriormente, a polivalência da escrita portuguesa é originalmente um acidente histórico: aqui ela decorre do fato que o /k/ latino, no caminho do português, tornou-se /ts/ e depois /s/ antes de vogal posterior (/e/, /ɛ/, /i/), fundindo-se

<sup>17</sup> Uma análise alternativa seria postular que a letra ⟨e⟩ final seleciona a variante “longa” da vogal da raiz, ou que letras distantes como ⟨a...e⟩ em ⟨sane⟩ são um dígrafo. Porém, mesmo por esta análise, continua sendo verdadeiro o fato que a escrita acaba representando algo morfológico: um dado morfema está associado a uma sequência de letras, independente da variação alomórfica da pronúncia. A situação do inglês também se deve a mudanças históricas na fonologia; no caso, a chamada Grande Mudança Vocálica (*Great Vowel Shift*), que distanciou significativamente o inglês escrito do falado (ROGERS, 2004, p. 191).

com o /s/. Porém, depois que a polivalência emergiu, a escrita aproveitou-se dela para conseguir dois significantes distintos onde a fala só tem um.<sup>18</sup> Quando uma sequência sonora tem vários sentidos em potencial, chamamos o fato de **homofonia** (“sons iguais”) ou de **polissemia** (“muitos sentidos”); e, se esses sentidos são distinguidos na escrita, falamos em **heterografia** (“escrita diferente”).

No exemplo 3.18, ⟨sem⟩ vs. ⟨cem⟩, as palavras são atualmente homófonas, mas elas descendem de termos originalmente distintos na fala: *sine* vs. *kentum*, respectivamente. Do ponto de vista histórico, diacrônico, poder-se-ia dizer que a ortografia manteve uma distinção que se perdeu na fala. Do ponto de vista do falante atual (sincrônico), o significante tem vários sentidos possíveis na fala, mas a escrita tem um recurso para especificar qual deles está em uso—ou seja, a representação grafêmica resolve uma ambiguidade presente na fonêmica.

Já no exemplo 3.20, ⟨viagem⟩ vs. ⟨viajem⟩ (substantivo “a viagem” vs. verbo “que elas viajem”), ambas as palavras derivam da mesma raiz de *viaticum*. Caso semelhante é ⟨mal⟩ (o contrário de “bem”) vs. ⟨mau⟩ (o contrário de “bom”), ambos da raiz latina *mal-*. Aqui, uma sequência fonológica que, até onde podemos saber, sempre foi homófona e polissêmica, teve o seu sentido tornado mais claro na escrita através da polivalência.<sup>19</sup>

#### 3.2.2.4 Símbolos e elementos não-fonográficos

Estivemos até o momento trabalhando com letras do alfabeto que representam fonemas, ou seja, fonogramas: sejam elas regulares, ou plurais/polivalentes. Mas os fonogramas não são os únicos tipos de grafema necessários em um texto português. Nesta seção, analisaremos outros tipos de signos.

<sup>18</sup> Podemos comparar esta cooptação com o conceito biológico de **exaptação**. É bem conhecido o princípio da “adaptação”, segundo o qual, para atender a pressões seletivas do ambiente, os seres vivos desenvolvem órgãos ou instintos em resposta a elas. Diferentemente disso, na exaptação um órgão ou instinto que surgira em resposta à determinada pressão é reaproveitado para uma função diferente. Por exemplo, acredita-se que as penas surgiram como reguladores de temperatura; mas, uma vez que estavam disponíveis, serviram à evolução do voo (GOULD; VRBA, 1982). Segundo Dehaene (2009), a própria escrita é uma exaptação que coopta o módulo cerebral de reconhecimento de arestas visuais, conectando-o ao módulo linguístico.

<sup>19</sup> Estamos enfatizando a tendência da escrita a buscar clareza de sentido; mas, evidentemente, a fala também se beneficia de clareza, e pode lançar mão de seus próprios recursos para atingi-la. Um desenvolvimento histórico paralelo à heterografia ocorreu por vezes na fala: são as chamadas duplas de palavras, ou *doublets* (WRIGHT, 1982, p. 23-). Considere o latim *pēnsāre*, cujo sentido original era “pesar, medir”; e que, metaforicamente, passou a significar também “pensar, ponderar”. Uma mudança fonológica comum na passagem do latim ao português foi a redução de /ns/ para /s/, o que criou nosso “pesar”; mas, ao contrário da maioria das palavras com /ns/, neste caso a língua falada optou por manter também a forma mais clássica “pensar”, especializando as nuances originais para significantes distintos. A geração de tais duplas talvez possa se inspirar conscientemente na escrita heterográfica, como quando alguns brasileiros deliberadamente pronunciam ⟨à⟩ mais longamente que ⟨a⟩; Wright, op.cit., menciona que locutores da БСС tomam o cuidado de distinguir artificialmente a pronúncia de *guerilla* e *gorilla*, que normalmente são homófonos. Não está claro se esses esforços para “ler como se escreve” têm o potencial de realmente serem incorporados à língua falada.

## 3.2.2.4.1 A letra ⟨h⟩

Embora seja enumerada no alfabeto junto das demais letras, o ⟨h⟩ é um caso especial; é um grafema que nunca é empregado seguindo o princípio fonográfico (ou seja, que nunca equivale a um fonema). O ⟨h⟩ originalmente representava uma consoante fricativa não-vozeada glotal, /h/—o mesmo som que a letra atualmente denota no inglês e japonês.<sup>20</sup> Porém, já no latim antigo esta consoante tornou-se “muda”—isto é, ela caiu da fala, e persistiu como uma notação exclusivamente heterográfica (HERMAN, 2000). Na ortografia portuguesa, ela mantém esta função heterográfica quando no começo ou fim de palavra: cp. ⟨era⟩ / ⟨hera⟩, ⟨a⟩ / ⟨ah⟩. Já em posição medial (no meio de palavra), o grafema foi tomado para compor dígrafos—ou seja, como uma notação para plurigrafia:<sup>21</sup>

Tabela 9 – Usos da letra ⟨h⟩ em português: Heterografia e plurigrafia

Som	Grafia	Uso
/ɛra/	⟨era⟩ ou ⟨hera⟩	Heterografia
/aʒa/	⟨aja⟩ ou ⟨haja⟩	Heterografia
/a/	⟨a⟩ ou ⟨ah⟩	Heterografia
/uŋa/	⟨unha⟩	Plurigrafia
/iʎa/	⟨ilha⟩	Plurigrafia
/ʃa/	⟨chá⟩	Plurigrafia

## 3.2.2.4.2 Pontuação, tipografia e sentido

O que representam os sinais de pontuação? À primeira vista, parece haver uma certa correspondência entre a pontuação na escrita, e a **entonação** na fala—ou seja, a variação na velocidade, altura (grave/agudo) e intensidade (volume) da voz. Por exemplo, o sinal ⟨?⟩ normalmente corresponde à entonação ascendente (cada vez mais aguda), própria das perguntas. Porém, observe novamente o exemplo 2.3, da página 45, reproduzido a seguir:

<sup>20</sup> Atualmente a letra ⟨h⟩ tem assumido este papel em estrangeirismos, como em *hostel* (=albergue). No português brasileiro, este som conta como um alofone do fonema /r/, o que torna a letra ⟨h⟩ uma alternativa ocasional à letra ⟨r⟩ para o mesmo fonema (um caso de polivalência da escrita). Notoriamente, a letra ⟨h⟩ é assim empregada em onomatopeias, como “haha” ou “ahá!”. Contudo, tais usos permanecem periféricos.

<sup>21</sup> O uso do ⟨h⟩ para formar dígrafos tem origem na prática romana de empregar a letra ⟨h⟩ para representar as consoantes aspiradas do grego antigo, cujo som era semelhante (para os ouvidos latinos) a uma sequência de duas consoantes: ⟨φ⟩ /p<sup>h</sup>/ > latim ⟨ph⟩, ⟨θ⟩ /t<sup>h</sup>/ > ⟨th⟩, ⟨χ⟩ /k<sup>h</sup>/ > ⟨ch⟩. Mudanças fonéticas abrandaram tais consoantes para fricativas simples, /f/, /θ/ e /x/; isso transformou pares de letras latinas, como ⟨ph⟩, em dígrafos que representavam uma só consoante, como /f/. Consequentemente, a letra letra ⟨h⟩ passou a ser tida por um marcador de dígrafos—papel que se difundiu em diversas línguas europeias. De fato, até a sequência /hw/ do inglês antigo, originalmente grafada ⟨hw⟩ (como em ⟨hwæt⟩), passou a ser grafada invertidamente como ⟨wh⟩ (⟨what⟩), provavelmente por influência do papel do ⟨h⟩ em outras ortografias.

- (2.3) a. O chefe mandou um recado; todos os funcionários voltaram para casa.  
 b. O chefe mandou um recado: todos os funcionários voltaram para casa.  
 c. O chefe mandou um recado—todos os funcionários voltaram para casa.

No item 2.3.a, a segunda frase, “todos os funcionários voltaram para casa”, não é o conteúdo do recado; já em 2.3.b, ela precisa ser; e em 2.3.c, ambos os sentidos são possíveis. A diferença de sentido é marcada pelos três sinais de pontuação, mas não é marcada na fala nem por entonação nem por pausas. Trata-se de um caso de heterografia, no sentido visto na seção anterior: uma sequência fonológica polissêmica tem seu sentido especificado na escrita.

Não é verdade que a pontuação seja uma transcrição da entonação e pausas; ela cumpre uma função semelhante, mas suas estruturas são distintas (o que se torna claro se compararmos a pontuação ortográfica com as transcrições linguísticas de pausas, usadas em análise da conversação). Os signos da pontuação são indícios de um sistema próprio da escrita para organizar a linguagem, com estruturas sintáticas que lhe são particulares, como demonstrado em detalhes por Nunberg (1990).

#### 3.2.2.4.3 Espaços, sinais diacríticos e heterografia

A escrita parece poder se apropriar facilmente de qualquer tipo de sinal visual para fins de clareza de sentido. A ortografia portuguesa normalmente usa espaços para separar palavras.<sup>22</sup> Porém, em ⟨por que⟩ vs. ⟨porque⟩, o espaço é usado para fins heterográficos—para diferenciar o “‘por que’ de pergunta” (como no inglês *why* ou japonês *dōshite*) do “‘porque’ de resposta” (como em *because* ou *dakara*). Caso semelhante é o do chamado “acento diferencial”. As marcas que são adicionadas a grafemas básicos, como os acentos gráficos do português ou o *dakuten* do japonês, são chamadas de **sinais diacríticos**<sup>23</sup> (pode-se dizer que o sinal diacrítico é um grafo “preso”, i.e. que depende de outro para se acoplar). Considere por um momento a função dos sinais de “acento agudo” e “circunflexo”. Na ortografia portuguesa, as letras ⟨e⟩ e ⟨o⟩ são polivalentes, podendo representar dois fonemas diferentes (compare “ele” vs. “elo”; “oco” vs. “oca”). Os sinais de agudo e circunflexo servem para marcar a sílaba tônica, e ao mesmo tempo determinam qual das duas leituras deve ser usada.<sup>24</sup>

<sup>22</sup> Embora existam algumas idiossincrasias convencionalizadas, como a diferença entre ⟨embaixo⟩ vs. ⟨em cima⟩.

<sup>23</sup> Do grego διακριτικός *dia-kritikós* “distintivo, diferenciador”.

<sup>24</sup> Há um pequeno número de palavras cuja ortografia inclui acento agudo, mas a leitura usa as vogais altas: “também”, “mantém”, “alguém” etc. Isso se deve em geral a distinções de pronúncia entre o português europeu e brasileiro.

Tabela 10 – Polivalência das letras ⟨e⟩ e ⟨o⟩ e acento gráfico

⟨e⟩	/e/, /ɛ/	“ele” /ele/, “elo” /elo/
⟨ê⟩	/e/	“revês” /re'ves/
⟨é⟩	/ɛ/	“revés” /re'ves/
⟨o⟩	/o/, /ɔ/	“oco” /oko/, “oca” /ɔka/
⟨ô⟩	/o/	“vovô” /vo'vo/
⟨ó⟩	/ɔ/	“vovó” /vo'vɔ/

O acento gráfico normalmente é usado em casos onde pronúncia padrão (*default*) da ortografia resultaria em outra leitura. Por exemplo, ⟨ônibus⟩ sem circunflexo, pela regra da ortografia, teria a sílaba tônica no /bus/ (como em “urubus”); ⟨lápiz⟩ no /pis/ (como em “funis”); ⟨é⟩ sem o sinal de acento agudo seria lido como “e”.

Porém, em palavras como ⟨tem⟩ (singular) vs. ⟨têm⟩ (plural), ⟨porque⟩ (conjunção) vs. ⟨porquê⟩ (substantivo) etc., a marca de acento gráfico (diacrítica) não causa alteração nenhuma na leitura da palavra; sua função aqui é, novamente, pura heterografia.<sup>25</sup> Particularmente importante na aquisição da norma escrita portuguesa é a distinção sintático-semântica entre ⟨a⟩ (artigo, preposição, pronome) vs. ⟨à⟩ (artigo + preposição; feminino de “ao”).

Quando um acento ou outro diacrítico é usado para selecionar uma de várias pronúncias possíveis, é chamado um **determinativo fonético**; e, quando seleciona um de vários sentidos possíveis, **determinativo semântico** (DEFRANCIS, 1984). Determinativos são cruciais em escritas morfográficas como o chinês, como veremos na próxima seção.

Em todos os casos de heterografia, o que vemos é que, consistentemente, a escrita apropriou-se da polivalência para diferenciar signos que, na fala, partilham significantes idênticos. Isso é consistente com a proposta de Vachek, Nunberg e outros, de que a escrita tem como objetivo primário a clareza de sentido, e não simplesmente a transcrição da fala.

#### 3.2.2.4.4 Algarismos, sinais matemáticos e outros recursos

A escrita possui algumas formas de representação que são específicas a ela, como recursos tipográficos (negrito, itálico, mudança e tamanho de fonte), e a organização espacial bidimensional (títulos, seções, quebras de linha, tabelas, notas de rodapé). De especial interesse para nós é a notação matemática, como os algarismos arábicos. Considere o papel do símbolo ⟨1⟩ nos seguintes exemplos:

<sup>25</sup> Deve-se notar que em português europeu a palavra ⟨têm⟩ é lida de forma diferenciada como [ˈtɛ̃jɐ̃] (< latim *tenent*). Nossas observações sobre acento diferencial em ⟨têm⟩ aplicam-se ao português brasileiro. Agradeço a Henrique Martins de Moraes por apontar a etimologia.

(3.21) 1 kiwi

(3.22) 1 maçã

(3.23) 1º lugar

(3.24) 1ª posição

(3.25) 100 maçãs

(3.26) A história em quadrinhos *100 Bullets*

No exemplo 3.21, o grafema ⟨1⟩ pode ser substituído pela palavra “um”; no 3.22, por “uma”. Nos exemplos 3.23 e 3.24, ele representa o morfema *primeir-* (selecionado por um determinativo especial ⟨<sup>a</sup>⟩, ⟨<sup>a</sup>⟩ que tem papel duplo: indicar o sufixo de gênero, *-o* ou *-a*, e também especificar a leitura “ordinal”, *primeir-*, do grafema ⟨1⟩). No exemplo 3.25, os três grafemas ⟨100⟩ precisam ser lidos de uma só vez como a palavra “cem”; sendo que os grafemas ⟨-00⟩ indicam que estamos contando centenas, e o ⟨1-⟩, que contamos uma só delas (se fosse um ⟨2⟩, seriam duas centenas = “duzentos”, etc.). O exemplo 3.26 é semelhante, exceto que o grafema ocorre no contexto de um estrangeirismo (um título de obra em inglês); embora os significantes sejam os mesmos, o leitor sabe que, propriamente, deve interpretá-lo na língua do contexto, resultando em *One Hundred Bullets*.

Está claro que símbolos matemáticos como ⟨1⟩, ⟨+⟩, ou ⟨%⟩ não representam fonemas, nem mesmo morfemas ou palavras. De fato, eles não apontam para unidades linguísticas, mas sim para idéias matemáticas—unidade, adição, porcentagem etc. Símbolos que se referem diretamente às “idéias”, ou mais precisamente ao significado, são chamados **semantográficos**, **semasiográficos** ou **ideográficos** (ROGERS, 2004, sec. 14.1). Além da escrita matemática, outros sistemas do tipo incluem a notação musical e os sinais de trânsito. No passado, já se argumentou que a escrita em caracteres chineses seria ideográfica; ou seja, que seus símbolos seriam uma espécie de escrita universal, capaz de representar diretamente as idéias, sem precisar fazer referência às palavras ou fonemas de uma língua (ver seção 3.3.4 à página 100). Tal proposta revelou-se falsa; os caracteres chineses demonstravelmente representam unidades linguísticas, como veremos a seguir.

### 3.3 Uma escrita morfográfica: chinês

Analisamos na seção anterior como funciona a escrita pelo som, usando como exemplo o português. Descobrimos que, em alguns casos, o português não se limita a representar o som, mas faz afinal referência ao sentido; como na ortografia profunda

(3.2.2.3.1), na heterografia (3.2.2.3.2), na pontuação (3.2.2.4.2), ou na notação numérica (3.2.2.4.4). Agora veremos em detalhes como funciona a escrita chinesa, cujos caracteres também constituem o componente *kanji* do japonês; e, embora sejam famosos como símbolos de sentido, veremos que eles fazem referência também ao som.

Considere as seguintes palavras da língua chinesa mandarim, representadas abaixo em três notações diferentes:

**Tabela 11 – Exemplos de caracteres chineses**

	Pronúncia	Sentido	Escrita em <i>pīnyīn</i>	Escrita em <i>hànzì</i>
1	/xə˥/	“rio”	<i>hé</i>	河
2	/xə˥/	“harmonioso”	<i>hé</i>	和
3	/kɔw˥/	“cachorro”	<i>gǒu</i>	狗
4	/tɕʰwan˥/	“cachorro”	<i>quǎn</i>	犬
5	/xə˥ . kʰow˥/	rio-boca “estuário, boca do rio”	<i>hékǒu</i>	河口
6	/kɔw˥ . ɕiə˥/	cachorro-sangue “melodramático”	<i>gǒuxiě</i>	狗血

A primeira coluna representa a pronúncia, usando o Alfabeto Fonético Internacional (IPA). Os fonemas estão agrupados em sílabas, seguidas por símbolos que indicam o **tom lexical** (isto é, a variação de altura, agudo/grave, que nas línguas chinesas é própria de cada morfema e pode distinguir palavras). A escrita *pīnyīn*, ilustrada na terceira coluna, é uma escrita chinesa baseada no alfabeto latino. Na última coluna temos as mesmas palavras grafadas em caracteres chineses tradicionais, ou *hànzì*,<sup>26</sup> cujo emprego no japonês é o objeto de estudo desta dissertação.

Os sons do chinês e a notação *pīnyīn* são descritos no apêndice C (p. 215); mas para nossa discussão, não é importante entender os detalhes da pronúncia. O ponto que nos interessa é a forma com que a sequência de sons /xə˥/, presente nas linhas 1, 2 e 5, sempre é grafada pelas letras *hé* em *pīnyīn*; onde a letra ⟨h⟩ representa a consoante /x/, a letra ⟨e⟩ a vogal /ə/, e o diacrítico ⟨˥⟩ o tom ascendente, /˥/. Disto se vê que a escrita *pīnyīn* é fonográfica, e segue o princípio alfabético que estudamos ante-

<sup>26</sup> 漢字 *hànzì* (leitura japonesa: *kanji*) significa literalmente “caracteres Han”. O nome da dinastia Han (206 BCE–220 CE) é usado como etnônimo na cultura chinesa, de forma que “caracteres chineses” é uma tradução idiomática. Porém, a despeito da etimologia literal de *hànzì*, os caracteres chineses antecedem esta dinastia; os registros arqueológicos mais antigos são da dinastia Shang (ca. 1200–1045 BCE), e mesmo nesses registros, o sistema já está em um estágio avançado, indicando que sua origem, hoje perdida, é ainda mais antiga.

riormente; ela marca fonemas com letras e tons com diacríticos.<sup>27</sup> De fato, a própria palavra *pīnyīn* 拼音 significa “sons soletrados” (DEFANCIS, 1984, p. 256).

Observe em particular que, nas linhas 1 e 5, a sequência *hé* representa o mesmo morfema, “rio”, enquanto que em 2 ela representa um morfema distinto, “boca” (ou seja, um homófono). Note que a grafia em *pīnyīn* não se altera; não há heterografia, ela depende apenas do som.

A última coluna grafa as mesmas palavras, mas usando os caracteres chineses clássicos. Novamente, os detalhes sobre os caracteres não são importantes no momento. O relevante é notar que o som *hé* é representado pelo mesmo caractere 河 nas linhas 1 e 5, onde ele significa “rio”; mas por um caractere distinto 和 na linha 2, onde o sentido é “harmonioso”. Isto é semelhante ao que vimos sobre a heterografia no português, como em “sem/cem” (cf. seção 3.2.2.3.2 à página 64 e seguintes); mas, enquanto a nossa heterografia é um fenômeno esporádico, criado a partir de acidentes como a polivalência gráfica, no sistema de escrita chinês ela é sistemática. Consistentemente, sentidos diferentes são grafados por caracteres diferentes.<sup>28</sup>

Compare agora as linhas 3 e 4. As palavras *gǒu* e *quǎn* são sinônimas; isto é, seu significante é diferente, mas seu significado é o mesmo (como as palavras portuguesas “cachorro” e “cão”). Observe que cada termo possui seu próprio caractere *hànzì*: 狗 vs. 犬. Ou seja, o caractere chinês não é um ideograma ou semantograma (seção 3.2.2.4.4); ele não representa somente o sentido, mas também o som. Se o significante muda (cf. linhas 3 e 4), ou se o significado muda (linhas 1 e 2), o caractere chinês também muda; ele representa integralmente os próprios signos da língua, tanto o som quanto o sentido.

Por fim, observe as palavras nas linhas 5 e 6. Trata-se de palavras compostas a partir de dois morfemas; e cada morfema corresponde a um caractere. Note que, em 河口 *hékǒu*, o sentido da palavra pode ser deduzido a partir dos sentidos dos morfemas: “rio-boca” = “boca do rio” = “estuário”. Em contraste, 狗血 *gǒuxuě* possui um sentido próprio, distinto dos sentidos de seus morfemas: “cachorro-sangue” = “sangue de cachorro” = “melodramático”. Dizemos que palavras como *hékǒu* são composições **transparentes**, e palavras como *gǒuxuě*, **opacas** (compare o português “paraquedas” vs. “amor-perfeito”).

Comparando as composições 5 e 6 com sua notação em *hànzì*, vemos que, mesmo

<sup>27</sup> Cf. definições desses termos na seção 3.2 acima

<sup>28</sup> DeFrancis (1984, p. 121) afirma que

O processo de desambiguação [=heterografia] em inglês geralmente resulta de mudanças históricas na pronúncia, como no conhecido caso de *knight* versus *night*. Em chinês, por outro lado, a técnica parece ter sido criada deliberadamente [cf. nossa tabela 18 à página 82, e discussão subsequente.]

que a palavra tenha um sentido opaco próprio, este sentido da palavra não é marcado de nenhuma forma nos caracteres. Ao invés disso, há sempre exatamente um caractere para cada morfema. Disto concluímos que os *hànzì* representam o sentido dos morfemas, não das palavras.

Reunindo o que vimos até agora, notamos que na escrita *hànzì*:

1. O número de caracteres equivale ao número de morfemas;
2. O mesmo morfema é sempre representado pelo mesmo caractere, qualquer que seja o sentido da palavra;
3. Morfemas homófonos (mesmo som, sentido diferente) são representados por caracteres distintos;
4. Morfemas sinônimos (mesmo sentido, som diferente) são representados por caracteres distintos.

Disto podemos concluir que, assim como as letras do alfabeto representam fonemas, os caracteres chineses representam morfemas: unidades de som e sentido. Chamamos este princípio de representação de **princípio morfográfico**, e os caracteres nele baseados de **morfogramas**. No chinês, como explica o morfologista Packard (2000, p. 307),

[...] a relação entre a forma ortográfica visual e a entrada lexical é virtualmente isomórfica, com cada forma visual (i.e. caractere) identificando univocamente uma entrada lexical (i.e. morfema).

Atualmente, a única escrita morfográfica em uso é a chinesa (cujos caracteres são empregados, com variantes distintas, na China, Taiwan e Japão). Há mais exemplos históricos, notoriamente o cuneiforme sumério, os hieróglifos egípcios e a escrita maia. Todas as escritas fonográficas conhecidas derivam direta ou indiretamente das morfográficas; e todas as escritas atualmente em uso, dos sistemas cuneiforme ou do chinês. A escrita portuguesa descende, em última análise, da cuneiforme (passando pela fenícia, grega e latina); a japonesa, do sistema chinês.

Somado ao fato que, no chinês, os morfemas normalmente são monossilábicos,<sup>29</sup> a adoção bastante regular do princípio morfográfico faz com que a seguinte equação quase sempre se aplique ao texto em chinês (FENG, 2009, p. 06, PACKARD, 2000, pp. 129-130):

(3.27)  $1 \text{ h\`{a}nz\`{i}} : 1 \text{ morfema} : 1 \text{ s\`{i}laba}$

<sup>29</sup> Os morfemas são em geral monossilábicos, mas as palavras não. Cf. seção 3.3.3.3 à página 90.

I.e., cada *hànzì* corresponde a exatamente um morfema e uma sílaba.<sup>30</sup>

### 3.3.1 Contagem dos caracteres chineses

Os maiores dicionários de caracteres chineses listam dezenas de milhares de caracteres distintos; o clássico *Kāngxī* 康熙 inclui 47.035, enquanto que o *Zhōnghuá Zìhǎi* 中华字海 alcança 85.568 (YUSHU; TINGJING et al., 1716; LENG; WEI et al., 1994). A grande maioria destes, porém, são caracteres muito raros, ou variantes históricas obscuras.<sup>31</sup> Além disso, variantes tratadas como caracteres distintos em alguns dicionários são consideradas alógrafos em outros.

O número de caracteres efetivamente usados pela população geral está na casa dos milhares. Para ter acesso a textos de nível médio, é esperado do estudante de ensino básico que, ao completar o 9º ano, tenha memorizado 3500 caracteres frequentes, que fazem parte do sistema de ensino (NPC, 2001). Um leitor adulto e assíduo talvez alcance o dobro desta cifra. Assim como as palavras do léxico linguístico, a frequência de uso dos caracteres segue uma curva logarítmica (ou seja, alguns poucos caracteres são enormemente mais frequentes que a grande maioria; ver capítulo 4).

### 3.3.2 Estrutura dos caracteres chineses

Considere o caso das “letras com acento” (ou seja, com diacrítico) no alfabeto português, como ⟨ã⟩. Por um lado, uma combinação dessas é um grafema, no sentido de ser uma unidade da escrita que representa um fonema. Por outro, ela não é uma grafema atômico, indivisível: ela pode ser decomposta em dois signos, a base ⟨a⟩ (indicando a vogal quanto à posição e abertura) e o diacrítico ⟨~⟩ (marcando a nasalidade).

Similarmente, os caracteres chineses podem ser atômicos, ou podem ser composições, formadas a partir de elementos visuais elementares que se recombinaem. Chama-

<sup>30</sup> Leitores familiares com a escrita japonesa devem atentar ao fato que a escrita chinesa é mais regular em três pontos:

- Um caractere quase sempre representa exatamente uma sílaba (ao contrário de, por exemplo, 政 = *matsurigoto*, no japonês);
- Um caractere quase sempre representa exatamente um morfema (ao contrário de, por exemplo, 唇 = *kuchi-biru*);
- E normalmente só há uma leitura possível (ao contrário de, por exemplo, 犬 = {*ken/inu*}).

Em outras palavras, a leitura chinesa dos caracteres apresenta muito menos *pluralidade* e *polivalência* que a japonesa (cf. seções 3.2.2.2–3.2.2.3, pp. 59–). Veja porém a seção 3.3.3 (p. 87) para exceções.

<sup>31</sup> Uma grande parte dos dicionários é dedicada a *hapax legomena*: caracteres que ocorrem uma só vez em todo o corpus da língua. Isso se dá porque, no passado, os escribas sentiam-se mais à vontade para criar os próprios caracteres, seguindo os princípios discutidos nesta seção. A maioria dos caracteres assim criados não se difundiu, tornando-se singularidades. Atualmente, a criação de caracteres não é mais prática comum.

mos estes elementos básicos de **componentes**. Ao contrário do alfabeto, a maioria dos caracteres chineses é não-atômica, ou seja, é composta por dois ou mais componentes. Compare os exemplos:

**Tabela 12 – Componentes dos caracteres chineses**

Morfema	Caractere	Componentes
1 <i>kǒu</i> “boca”	口	口
2 <i>hé</i> “harmonioso”	和	禾 口
3 <i>xíng</i> “forma”	形	开 彡
4 <i>xíng</i> “camiseta”	衫	衤 彡
6 <i>rán</i> “queimar”	燃	火 然 → 月 犬 灬
7 <i>hé</i> “rio”	河	氵 可
8 <i>gǒu</i> “cachorro”	狗	犭 句

Note que:

- O exemplo 1, 口 *kǒu* “boca”, é um caractere indivisível; os demais são composições, formados por vários componentes.
- Este mesmo caractere 口 aparece como componente no exemplo 2, 和 *hé* “harmonioso”. Componentes podem ser livres, i.e. aparecer também como caracteres independentes, como é o caso de 口 *kǒu*; ou podem ser elementos presos, que só aparecem em combinações. O elemento 彡, comum a 3 e 4, 衫 e 形, é um componente preso; ele nunca é usado sozinho. A maioria dos componentes é livre. Perceba que um componente livre também é um caractere, o que implica que ele corresponde a um morfema (e portanto possui um som e um sentido).
- O exemplo 6, 燃 *rán* “queimar”, possui um componente à direita, 然 *rán* “desta forma”. Este componente 然, por sua vez, é composto ele próprio por três componentes, 月 犬 灬, dispostos em um triângulo. O processo de composição visual é recursivo; isto é, pode ser reaplicado aos próprios caracteres compostos, gerando composições de múltiplos níveis.<sup>32</sup> Além disso, a composição permite diversas configurações espaciais, como direita/esquerda, triângulo, cima-baixo etc. Não é possível variar a disposição; uma vez que um caractere esteja estabelecido, seus componentes sempre precisam ser grafados na mesma posição.

<sup>32</sup> Palavras como “recursivo” e “preso” têm sentido técnico na linguística, referendo-se a características da linguagem. Porém, usamo-las aqui como metáforas para descrever padrões puramente visuais, não-linguísticos. Estas metáforas funcionam bem; os padrões de composição seguem certas regras implícitas que são análogas às regras da fonologia ou morfo-sintaxe, tais como a recursividade. Myers (2015) discute mais exemplos e realiza testes de aceitabilidade de algumas regras dessa “gramática visual” do caractere chinês.

- Nos exemplos 6, 7 e 8, 燃, 河 e 狗, os componentes 灬, 氵 e 犳 são considerados formas alternativas (alógrafos) de caracteres livres: 氵 de 水 *shuǐ* “água”; 犳 de 犬 *quǎn* “cão”; e 灬 de 火 “fogo”. Isto mostra como os componentes podem apresentar variação alográfica, condicionada por sua posição (à esquerda, embaixo...)

### 3.3.2.1 Contagem dos componentes

Assim como o número de caracteres, também não é simples precisar quantos componentes existem, pois a contagem depende dos critérios escolhidos para determinar o que conta como “o mesmo componente”—e também de quantos caracteres obscuros são inclusos no corpus. O número na prática parece ser pouco mais de 1000; [Wieger \(1927\)](#) conta 1027 e [Karlgren \(1923\)](#), 1350. Segundo nossa estimativa, o projeto KanjiVG atualmente usa cerca de 1258 componentes para compor 6412 caracteres frequentes no japonês ([Apel et al. \(2013\)](#); cf. também capítulo 4 à página 131). Muitos componentes dessas contagens não são elementares graficamente (isto é, são compostos por outros componentes), mas foram tratados como unidades distintas devido ao fato de possuírem sentidos próprios.

A educação no Japão e China é baseada na unidade “caractere”; embora os componentes internos sejam mencionados, eles não são listados, memorizados ou praticados individualmente. Porém, há associações interessantes entre som, sentido, e os componentes; e, ainda que essas associações não sejam analisadas conscientemente, há evidência de que elas são internalizadas pelos aprendizes à medida que avançam no processo de letramento ([CHEN et al., 2003](#)). As associações podem ser semânticas, fonológicas, ou ambas. Consideremo-las uma a uma.

### 3.3.2.2 Componentes semânticos

Considere os seguintes caracteres e os morfemas que representam:

**Tabela 13 – Componentes semânticos**

1	狗	<i>gǒu</i> “cachorro”
2	猫	<i>māo</i> “gato”
3	猿	<i>yuán</i> “macaco”
4	狩	<i>shòu</i> “caçar”
5	犯	<i>fàn</i> “crime”
6	狎	<i>xiá</i> “íntimo”
7	猶	<i>yóu</i> “bem como, ainda”

Os exemplos 1, 2 e 3 são nomes de animais, e todos incluem o componente 犭 à esquerda. 4, 狩 *shòu* “caçar”, inclui o mesmo componente e, embora não signifique um animal, refere-se a uma ação relacionada a eles. Desta forma, o componente 犭 age como uma pista que indica, de forma algo vaga, o campo semântico do morfema representado pelo caractere. Dizemos que 犭 é um **componente semântico** (意符 *yìfú*, japonês *ifu*). Como mencionado anteriormente, 犭 é considerado alógrafo de 犬 *quǎn* “cão”; mas quando usado como componente, ele indica o campo semântico de “animal”.<sup>33</sup>

Componentes semânticos não são indicadores perfeitos. 5, 犯 *fàn* “crime”, talvez esteja classificando ações criminosas como animais, e um leitor mais imaginativo talvez possa encontrar várias justificativas para 6 狎 *xiá* “íntimo”; mas um morfema mais abstrato ou gramatical, como 7 猶 *yóu* “ainda”, é difícil de associar semanticamente exceto arbitrariamente. Não há nenhum registro do raciocínio original de quem compôs os caracteres, e o sistema de ensino não se ocupa explicitamente dessas interpretações (o que significa que nenhuma análise é oficializada como canônica). Note porém que mesmo que o componente semântico não aja como pista do campo semântico, ainda assim ele age como uma forma gráfica de distinguir sentidos diferentes com o mesmo som, exatamente como a ortografia heterográfica do português (e.g. ⟨sem⟩ vs. ⟨cem⟩). Vimos acima, na seção 3.2.2.4, que um elemento gráfico acrescentado para determinar o sentido é chamado **determinativo semântico**. Na escrita chinesa, bem como no sumério, egípcio e maia, os determinativos são uma peça fundamental do sistema. Em revisão da literatura psicolinguística, Kess e Miyamoto (1999) (p. 201) afirmam que caracteres carregando os mesmos componentes são agrupados mentalmente em “famílias”, cuja ativação cerebral é conjunta.

Vimos até agora duas funções do componente semântico: originalmente, eles indicam o campo semântico do morfema; em alguns casos essa indicação é imperfeita, mas ainda assim serve para distinguir o sentido, como um determinativo. Há ainda uma terceira função possível. Alguns componentes semânticos foram selecionados como chaves para busca em dicionários; são os chamados “cabeçalhos de seção” 部首 (*bùshǒu*, jap. *bushu*). O conjunto exato de componentes assim empregado depende do dicionário; o mais importante é o conjunto de 214 cabeçalhos do dicionário *Kāngxī*. Por exemplo, *quǎn* 犭 / 犬 acima é usado como o cabeçalho número 94 neste dicionário.

As expressões “cabeçalho de seção” *bùshǒu* e “componente semântico” *yìfú* foram historicamente traduzida no Ocidente como “radical”, comparando-os com os radicais da etimologia—uma escolha infeliz, visto que os componentes semânticos, além de vagos, em geral são determinativos, isto é, acréscimos tardios aos caracteres (ou seja,

<sup>33</sup> Mais especificamente, este componente é usado para animais peludos (bestas ou mamíferos). Há outros componentes semânticos para pássaros (鳥、隹), peixes (魚), insetos e répteis (虫), etc.

o *contrário* dos radicais, que são as bases às quais se acrescentam afixos). Além da analogia inválida, a tradução “radical” mistura os conceitos de componente semântico e cabeçalho de seção. É importante distingui-los: muitos componentes semânticos não são listados como cabeçalhos—e muitos cabeçalhos não têm papel semântico, são outros tipos de componentes.

Considere novamente o caractere 狗 *gǒu*, “cachorro”. Vimos na tabela acima que o componente à esquerda, 犭 “animal”, indica o campo semântico, como fica claro ao comparar com os demais caracteres que o compartilham. E quanto ao componente à direita, 句; qual seria o seu papel? É o assunto da próxima seção.

### 3.3.2.3 Componentes fonéticos

Compare a pronúncia dos seguintes morfemas, e o elemento visual em comum nos caracteres:

**Tabela 14 – Componentes fonéticos**

1	狗	<i>gǒu</i> “cachorro”
2	苟	<i>gǒu</i> “supondo que”
3	枸	<i>gǒu</i> “cidra” (planta)
4	鉤	<i>gōu</i> “gancho”
5	佢	<i>kòu</i> “raquítico, corcunda”
6	句	<i>jù</i> “frase, verso”
7	駒	<i>jū</i> “potro”
8	拘	<i>jū</i> “capturar”
9	洵	<i>xù</i> “aquecido”

Os exemplos 1 a 3, 狗, 苟 e 枸, todos compartilham o componente 句, e todos são pronunciados de forma idêntica: *gǒu*. Além disso, os exemplos 4 鉤 e 5 佢 são pronunciados *gōu* e *kòu*, que são pronúncias bastante semelhantes; só o que muda é o tom, e (no caso de *kòu*) a aspiração da consoante, de ⟨g⟩ = /k/ para ⟨k⟩ = /kʰ/. Dizemos que 句 aqui age como **componente fonético** (音符 *yīnfú*; jap. *onpu*)—indicadores da pronúncia do caractere.<sup>34</sup> O caso de 4 e 5 mostra que os componentes fonéticos não são exatos; assim como os semânticos, muitas vezes eles servem apenas como uma pista geral, suficiente para que um leitor nativo consiga recuperar da memória o item lexical relevante. Além disso, um componente fonético pode ter mais de um grupo de leituras atribuído

<sup>34</sup> Mais precisamente, seriam componentes *fonológicos*. Estamos seguindo a terminologia já estabelecida por DeFrancis (1984) e outros autores da área; “fonético” aqui não se refere a “fones” mas sim, de forma ampla, a “som, leitura, pronúncia”.

a ele, como mostram os exemplos 6 a 9.<sup>35</sup> Poderíamos parafrasear o sinal enviado pela presença de 句 como: “A pronúncia deste caractere provavelmente é semelhante ou a *gǒu*, ou a *jù*” (isto é um caso de polivalência de leitura, como o ⟨c⟩ do português, que pode representar /k/ ou /s/).

Combinando o que vimos na seção anterior sobre 犭 ao que vimos nesta sobre 句, podemos entender a estrutura do caractere 狗, que poderia ser descrita assim:

- 犭 : Este caractere representa um morfema relacionado a animais peludos
- + 句: A pronúncia é próxima de 句 {*gǒu* ou *jù*}
- = 狗: O morfema *gǒu* “cachorro”.

狗 é uma juxtaposição onde um componente age como pista de sentido, e outro, como pista de pronúncia. Sem esses “ganchos” mnemônicos, provavelmente seria impossível dar conta da memorização de milhares de caracteres.

#### 3.3.2.4 Tipos de caracteres chineses: A classificação tradicional

Vimos que os caracteres chineses podem ser simples (ou atômicos), isto é, podem ser formados por um só componente; ou então podem ser compostos, juxtapondo dois ou mais deles. O que cada componente simboliza, e qual a relação entre eles? Isso depende do tipo do caractere; existem vários princípios de construção diferentes, gerando categorias distintas. No campo de estudo dessa tipologia, o pioneiro e grande clássico é o dicionário *Shuōwén Jiězì* 說文解字, escrito por Xu Shen no século II (O’NEILL, 2013).<sup>36</sup> Devido à sua antiguidade, é tentador atribuir uma posição privilegiada às teorias de Xu Shen, como se o autor estivesse próximo da fonte original dos *hànzì*. Porém, é

<sup>35</sup> Note que a letra ⟨j⟩ em *pīnyīn* representa /tɕ/, enquanto ⟨x⟩ representa o som semelhante /ç/, de forma que 9 soa bastante próximo de 6–8.

<sup>36</sup> O título traduz-se como “Explicando *Wén* 文; Analisando *Zì* 字”. Os termos *wén* e *zì* ambos se referem à escrita; as duas palavras não parecem ter sido distinguidas claramente antes do *Shuōwén*, e a distinção que Xu Shen faz é pouco clara, sendo objeto de discussões até hoje. Shen escreveu no Posfácio do dicionário que os *zì* nascem ou são “filhos” dos *wén*—como argumento, ele destaca a presença do componente 子 “filho” no grafema *zì* 字. E isso é tudo que o autor nos deixou. Destacamos duas interpretações desta afirmação críptica:

- A interpretação tradicional postula que a palavra *wén* se referiria a grafemas atômicos ou indivisíveis, e *zì* a grafemas compostos, criados a partir da combinação dos simples. Então os *zì* “nasceriam” dos *wén* porque aqueles são criados a partir da juxtaposição destes.
- Porém, Bottéro (2004) e O’Neill (2013) demonstram que em nenhuma passagem do texto do *Shuōwén* o autor faz uma tal distinção *wén/zì* em termos de grafemas simples vs. compostos—nem mesmo nas definições que o próprio dicionário faz das duas palavras. Os autores consideram todas as passagens nas quais Xu Shen emprega os dois termos, e concluem que o filólogo consistentemente usa *wén* para se referir aos caracteres do ponto de vista da estrutura visual, grafológica; a principal preocupação do autor ao falar de *wén* é discutir quais são os componentes do caractere, e como a identidade desses componentes foi alterada historicamente. Já a palavra *zì* é usada por Shen para discutir os caracteres enquanto representação da língua, de morfemas, incluindo sua pronúncia e sentido.

importante ressaltar que Shen viveu mais de um milênio depois da formação da escrita chinesa, de forma que suas análises são hipóteses de reconstrução histórica; muitas delas estão hoje descredenciadas, devido à descoberta recente de materiais arqueológicos aos quais Xu Shen não pôde ter acesso. Porém, as análises de Shen são muito importantes culturalmente, no sentido de que elas sustentaram o peso da autoridade filológica por quase dois mil anos, sendo citadas em dicionários, material educativo etc.

Este trabalho visa a análise sincrônica, i.e. do sistema de escrita tal como ele é hoje, do ponto de vista do leitor moderno, sem usar de informações históricas que só um filólogo saberia. Contudo, a melhor forma de entender a tipologia dos caracteres é rever brevemente o processo formativo de cada categoria. Ao dicionário de Xu Shen é atribuída a difusão de um modelo que classifica-as em seis tipos: as chamadas “seis [formas de] escrita”, *liùshū* 六書 (japonês *rokusho*). Os termos apresentados por Shen permanecem em uso até hoje, tanto na educação quanto na terminologia linguística, então é interessante detalhá-los a um.<sup>37</sup>

O **pictograma**, 象形 *xiàngxíng* (literalmente “representando a forma”; jap. *shōkei*), é originalmente um desenho: suas linhas gráficas imitam os contornos de um objeto externo, ele “curva-se e dobra-se de acordo com a forma [do objeto]”. Poucos caracteres chineses são pictogramas; porém, embora a quantidade seja baixa, muitos pictogramas são caracteres de alta frequência. Além disso, eles são bastante empregados como componentes em caracteres compostos.

À medida que a escrita chinesa foi convencionalizada, padronizada e simplificada para fins de caligrafia e tipografia, a iconicidade dos pictogramas diminuiu—isto é, os desenhos originais foram se tornando símbolos mais opacos, abstratos.<sup>38</sup>

---

Então, segundo as análises de Bottéro e O’Neill, dizer que *zì* nasce de *wén* equivale a dizer que “o sentido linguístico é determinado pela estrutura grafológica”. Na cosmologia da época, acreditava-se que a escrita havia sido criada pelo sábio mitológico Cangjie inspirando-se nos “padrões” (*wén*) da natureza, e.g. as pegadas deixadas por pássaros. Então os “padrões” gráficos dos caracteres (*wén*) refletiriam um sentido original dado a eles pelos sábios do passado. Qualquer mudança na estrutura equivaleria a uma corrupção, e o projeto do dicionário de Xu Shen seria corrigir essas distorções. Portanto, segundo estas análises, *Shuōwén Jiězì* significaria algo como “explicando a estrutura gráfica dos caracteres [e sua evolução histórica] para analisar [e recuperar o sentido original de] seu uso linguístico”.

<sup>37</sup> Seguimos aqui a exposição de Boltz (1994), pp. 143–.

<sup>38</sup> Imagens das formas antigas extraídas de Sears (2013). Tabelas evolutivas como esta são comuns ao discutir pictogramas chineses, mas é importante ter em mente que a progressão não foi uma simples sequência linear; por séculos, muitas variantes coexistiram em contextos diferentes, e em alguns casos apareceram novas variantes até mais realisticamente pictográficas que as primitivas. O sistema moderno usa grafos bastante padronizados, mas na China antiga a diversidade visual era grande. As formas “selo” ainda hoje são usadas em carimbos de identificação pessoal (donde o nome).

Tabela 15 – Progressão histórica dos pictogramas chineses (*xiàngxíng*)

Fase do sistema Época	Ossos oraculares ca. séculos XII–X AC	Selo ca. sec. II AC	Moderno ca. sec. II DC
“cão”			
“cavalo”			
“montanha”			

Parece razoável supor que os pictogramas tenham sido originalmente símbolos não-linguísticos, referindo-se aos conceitos diretamente; isto é, que o pictograma de “cavalo”, em um primeiro momento, representava o conceito de “cavalo” em si e não a palavra “cavalo”. Um tal sistema de símbolos é denominado **proto-escrita** ou **predecessor da escrita**. Comparado à escrita propriamente, uma proto-escrita baseada na pictografia é bastante limitada: ela só pode representar objetos tangíveis, desenháveis, e não tem uma maneira de representar univocamente todos os enunciados da língua. A escrita propriamente começa quando um grafema como 馬 passa a representar não simplesmente a idéia “cavalo”, mas sim a palavra *mǎ* (no chinês antigo \**mrâ* ).<sup>39</sup>

Outra categoria tradicional é o caractere **indicativo** ou **analógico**, 指事 *zhǐshì* (literalmente “indicando as coisas”; jap. *shiji*). Observe:

<sup>39</sup> Nossos exemplos de chinês antigo seguem a reconstrução e notação de [Schuessler \(2009\)](#), q.v.

Tabela 16 – Caracteres chineses indicativos ou analógicos (*zhǐ shì*)

	Forma antiga	Forma atual	Analogia
“Cima”			Uma linha horizontal curta <b>acima</b> de uma linha de referência
“Baixo”			Uma linha horizontal/diagonal curta <b>abaixo</b> de uma linha de referência
“Três”			Três linhas denotam o número “três”
“Raiz, origem”			Uma linha horizontal curta aponta a raiz no pictograma “árvore” 木

Nos caracteres indicativos, o significado é apontado por meio de uma analogia visual; por exemplo, um traço que destaca parte de um pictograma, ou uma relação geométrica que remete à palavra representada.

Os caracteres de **associação semântica**, 會意 *huìyì* (lit. “combinando sentidos”; jap. *kai'i*), sugerem um morfema através da juxtaposição associativa de dois ou mais componentes:

Tabela 17 – Caracteres chineses de associação semântica (*huìyì*)

Sentido	Caractere	Associação
“Bosque”		Árvore 木 + árvore 木 = bosque
“Floresta”		Árvore 木 ×3 = floresta
“Descansar”		Pessoa 亻 ao lado de árvore 木 = descansar
“Colheita”		Mão 采 acima de árvore 木 = colheita

Tal “álgebra dos sentidos” há muito chama a atenção de observadores ocidentais, tais como Leibniz ou Fenollosa, que acabaram exagerando sua importância para a escrita chinesa. Discutiremos este ponto na seção 3.3.4 (p. 100).

Os três tipos de caracteres vistos até agora (pictogramas, indicativos e associações semânticas) só permitem representar com naturalidade alguns tipos de morfemas.

É difícil usar estas técnicas visuais para indicar morfemas mais abstratos, como a ação de “vir” ou a direção “norte”, ou ainda morfemas mais gramaticais, como o pronome demonstrativo *zhī*. A solução encontrada foi o **princípio do rebus**, ou empréstimo homofônico (假借 *jiǎjiè*, jap. *kashaku*; lit. “empréstimo, (re)aproveitamento”). O rebus consiste em reaproveitar um caractere para representar um outro morfema de mesmo som (ou, com mais frequência, de som semelhante). Em nossa cultura contemporânea, esta técnica aparece como uma brincadeira em revistas de passatempo. Um exemplo seria:<sup>40</sup>

**Figura 4 – Exemplo de rebus (empréstimo homofônico de pictogramas) em português. Que frase está representada abaixo?**



A frase é: “Às sextas-feiras como caramelos”. Algumas sequências sonoras foram substituídas por pictogramas que representam palavras homófonas. “Rebus” é o nome que se dá a este tipo de símbolo que “pivoteia” pelo som. O termo vem do latim, onde é o ablativo de *rēs* “coisa”; ou seja, a representação do som “por meio de coisas”.

Semelhantemente, no chinês antigo:

**Tabela 18 – Caracteres chineses de rebus ou empréstimos homofônicos (*jiǎjiè*)**

Caractere	Uso original	Uso rebus
屮 之	*/tə/ > <i>zhī</i> = um tipo de planta	*/tə/ > <i>zhī</i> = pronome demonstrativo
𠂆 北	*/pâkh/ > <i>bèi</i> = costas	*/pâk/ > <i>běi</i> = norte
來 來	*/mrêk/ > <i>mài</i> = trigo, cevada	*/rêk/ > <i>lái</i> = vir

Podemos dizer que, para representar *zhī* “pronome demonstrativo”, o significado “planta” foi descartado do morfograma pictográfico 之, reduzindo-o a um fonograma que representa os fonemas *zhī*.<sup>41</sup> Convencionalizou-se então empregar esse fonograma para representar o pronome *zhī*, criando um novo morfograma, não mais um pictograma simples, mas um empréstimo homofônico. Processo análogo se deu com os demais caracteres rebus.

<sup>40</sup> Imagens extraídas de <<http://openclipart.org>> (OPENCLIPART.ORG, 2016), e modificadas pelo autor. Artistas originais: usuários casino, dannya, egore911, Tavin e PrinterKiller.

<sup>41</sup> Compare com a criação da escrita *kana* no Japão (3.5.5 à página 115), ou com a técnica *ateji* (3.5.3 à página 107).

O rebus foi uma invenção crucial, e não só na China; ele foi a chave que permitiu passar da proto-escrita à escrita plena, capaz de representar qualquer enunciado. Sistemas fonológicos como o nosso alfabeto são descendentes de rebus. Por exemplo, a letra ⟨M⟩ descende de um pictograma hieróglifo egípcio, , que originalmente representava “água”.<sup>42</sup> Ele foi tomado por rebus para representar o som /n/, baseado em uma antiga palavra egípcia para “água”, \**net* (BRIGHT; DANIELS, 1996, sec. 4; WARREN, 2005). Posteriormente, o mesmo símbolo foi usado pelos fenícios para representar o som /m/, devido ao fato que “água” em sua língua era pronunciado *mem*. Da escrita fenícia ele passou com esse valor para os alfabetos grego: ⟨M,μ⟩ e latino: ⟨M,m⟩.

Os caracteres **fono-semânticos**, 形聲 *xíngshēng* (literalmente “forma e som”; jap. *keisei*) possuem um elemento que indica a pronúncia, e outro que indica o campo semântico ou sentido geral. Os exemplos que vimos na seção sobre componentes fonéticos (3.3.2.3 à página 77) são normalmente deste tipo, como reproduzido abaixo:

Tabela 19 – Caracteres fono-semânticos (*xíng shēng*)

	Morfema representado	Componente	
		Semântico	Fonético
駒	<i>jū</i> “potro”	馬 = “cavalo”	句 = <i>jù</i>
拘	<i>jū</i> “capturar”	扌 = “mão”	句 = <i>jù</i>
晌	<i>xù</i> “aquecido”	日 = “sol”	句 = <i>jù</i>

Nos três exemplos acima, o componente semântico é (originalmente) um pictograma, enquanto que o componente fonético, 句, é um caractere que está sendo usado pelo som, ou seja, um rebus. Essa estrutura resulta de um processo histórico, que parte do fonético como base e acrescenta os semânticos (BOLTZ, 1994, pp. 59–). O desenvolvimento típico é como se segue:

1. No primeiro estágio, um caractere representa um único morfema—seja como pictograma, indicativo, ou associação semântica;
2. Depois, o caractere passa a ser usado como rebus para representar vários outros morfemas de pronúncia semelhante, criando ambiguidade (polivalência de sentido);
3. Finalmente, componentes semânticos são adicionados como determinativos, a fim de distinguir cada um dos morfemas homófonos (heterografia).

<sup>42</sup> O mesmo símbolo já era usado para marcar jarros d’água antes mesmo da invenção da escrita (NICHOLS, 2008). À época dos hieróglifos, o sinal já havia se estabelecido como fonograma; mas ainda podia ser usado para denotar o sentido de “água”, quando triplicado.

Consideremos o exemplo do rebus 北 *bèi*, citado acima, e do caractere fonosemântico que ele gerou, 背 *bèi* “costas”:

**Tabela 20 – Processo gerativo típico de um caractere fonosemântico**

Fase	Morfemas representados	Explicação
1	北 = <i>bèi</i> < *pâkh “costas”	Pictograma de duas pessoas de costas.
2	北 = <i>bèi</i> < *pâkh “costas”, 北 = <i>běi</i> < *pâk “norte”	Por rebus, o caractere passou a poder representar também <i>běi</i> “norte”. Neste estágio há ambiguidade, ou seja, polivalência de sentido: em cada ocorrência de 北, é preciso decidir qual dos dois morfemas está grafado.
3	背 = <i>bèi</i> < *pâkh “costas” 北 = <i>běi</i> < *pâk “norte”	Acrescentou-se o componente semântico “parte do corpo”, 月, para representar o morfema “costas”: 背. Com isso, 北 ficou reservado não-ambiguamente para “norte”.

No final deste processo, terminamos com um caractere composto (背 “costas”) que juxtapõe dois componentes: um que sugere o som (北 = *bei*), e outro para o sentido (月 = “parte do corpo”). Estes são os compostos fonosemânticos.

A categoria fonosemântica é a mais numerosa; um dado frequentemente citado é que mais de 90% dos caracteres chineses são deste tipo (DEFRANCIS, 1984, p. 96). Contudo, tal proporção só se aplica à “contagem de tipos” e não de “ocorrências”; ou seja, ela é uma contagem de dicionário, que não leva em conta o fato que muitos caracteres de alta frequência pertencem a outras categorias. Em um texto natural, as proporções serão diferentes.<sup>43</sup> Além disso, mesmo que um caractere seja fonosemântico em sua formação, nem sempre o papel fonético ou semântico dos componentes se manteve no sistema moderno (ou seja, é preciso distinguir diacronia de sincronia, origem de uso). Discutiremos estes pontos em mais detalhes no capítulo 4 (p. 131), que também apresentará estatísticas mais representativas. Qualquer que seja a medida, porém, é indiscutível que a categoria fonosemântica é numericamente predominante no chinês.

A última categoria listada por Xu Shen são os caracteres **redirecionados**, 轉注 *zhuǎnzhù* (jap. *tenchū*; literalmente “desviados-redirecionados”). Esta categoria é obscura e até hoje pouco compreendida. Na única menção que Shen faz do termo, ele cita como exemplo 考 *kǎo* “examinar” e 老 *lǎo* “idoso”, que são morfemas etimologicamente relacionados. Isso leva Boltz a fazer a seguinte tradução/interpretação da passagem

<sup>43</sup> Por exemplo, os caracteres simples, de um só componente (que nunca são fonosemânticos) são especialmente frequentes; no caso do japonês um teste contabilizou 64 caracteres dentre os 200 mais usados são simples, e ainda que 75% dos simples estão listados dentre os 1000 mais frequentes (HABEIN, 2000, p. 17).

(BOLTZ, 1994, p.144–145):

*Zhuǎnzhù*: Agrupam-se [vários grafemas de] categorias [semânticas similares] sob uma entrada, sendo que eles relacionam-se uns aos outros por sua etimologia em comum.

Devido à dificuldade de interpretação, a categoria *zhuǎnzhù* em geral não é mais usada—a despeito da difusão ainda ampla do termo “seis escritas”, 六書 *liùshū*, os materiais didáticos e lexicográficos atuais focam-se nas outras cinco.

Considerando as seis categorias como um todo, percebemos que quatro delas descrevem a lógica estrutural dos caracteres em si: eles podem ser pictogramas (desenhos), indicativos (analogias visuais), associações semânticas (juxtaposições de sentidos), ou fono-semânticos (juxtaposições de sentido e pronúncia). As outras duas categorias não se referem à estrutura dos caracteres, mas sim ao seu uso: empréstimos homofônicos (rebus) e redirecionamentos.

### 3.3.2.5 Componentes de papel duplo: os caracteres fonético-associativos duais

Vimos na seção anterior que o tipo mais frequente de caractere é o fono-semântico, que combina um componente semântico (que indica o sentido) a um fonético (que indica o som). Porém, em muitos casos, o componente fonético também serve simultaneamente como uma segunda pista de sentido. Observe:

	Som	Sentido	Componentes
1	包 <i>bāo</i>	“embrulhar”	
2	抱 <i>bào</i>	“abraçar”	手 <i>shǒu</i> “mão” + 包 <i>bāo</i> “embrulhar”
3	泡 <i>pào</i>	“bolha”	氵 <i>shuǐ</i> “água” + 包 <i>bāo</i> “embrulhar”
4	胞 <i>bāo</i>	“placenta”	肉 <i>ròu</i> “corpo” + 包 <i>bāo</i> “embrulhar”
5	雹 <i>báo</i>	“granizo”	雨 <i>yǔ</i> “chuva” + 包 <i>bāo</i> “embrulhar”
6	咆 <i>páo</i>	“rugir”	口 <i>kǒu</i> “boca” + 包 <i>bāo</i> “embrulhar”
7	庖 <i>páo</i>	“cozinha”	广 “prédio” + 包 <i>bāo</i> “embrulhar”

**Tabela 21 – Componentes fono-semânticos**

Em todos os exemplos acima, o componente 包 *bāo* tem valor como pista fonética (exata ou aproximada). Mas, além disso, em 2–4, o sentido “embrulhar” de 包 claramente combina-se com o sentido do outro componente, na mesma “álgebra dos sentidos” dos caracteres associativos: embrulho + mão = abraçar, embrulho + água = bolha, etc.

Já nos casos 5–7, a lógica semântica é mais discutível. Seria “granizo”, um “pacote de chuva”? Ou, nesse caractere, o componente 包 tem valor puramente fonético? E em 庖 “cozinha”? Com imaginação o suficiente, sempre é possível postular *alguma* associação semântica, mas não é simples decidir quais delas são processadas enquanto tais pelo leitor típico.<sup>44</sup> Zhou e Marslen-Wilson (1999) mostram experimentalmente que, ao processar componentes fonéticos, seu conteúdo semântico também é ativado no cérebro; parece plausível que esse conteúdo seja usado como pista, onde couber.

Em termos da classificação tradicional, este tipo de caractere pode ser considerado um fono-semântico 形聲 *xíngshēng* (jap. 形声 *keisei*) que é, ao mesmo tempo, também uma associação semântica 會意 *huìyì* (jap. 会意 *kai'i*). Tōdō (1965) denomina-os caracteres “fonético-associativos” (會意形聲 *ka'i-keisei*; mandarim 會意形聲 *huìyì-xíngshēng*); eles são o tipo fundamental em sua teoria dos caracteres chineses, que postula que famílias etimológicas de palavras estão refletidas em famílias de caracteres, marcadas pela presença do mesmo componente fono-semântico.

Para enfatizar o papel duplo que desempenham, nós chamaremos um componente que pode ter os dois papéis de **componente fonético-semântico dual**, e os caracteres identificados por TŌDŌ de **caracteres fonético-associativos**.

### 3.3.2.6 Limitações no papel dos componentes

Na discussão até agora, insistimos na palavra “pista” para enfatizar que as associações, tanto as fonéticas quanto as semânticas, são aproximadas e variáveis. Fontes de variação incluem os fatos que:

- Nem sempre está claro quais são os subcomponentes unitários de um caractere.
- A presença de um componente não garante som ou sentido; muitas leituras e significados são idiossincráticos, imprevisíveis a partir dos componentes visuais.
- As pistas semânticas podem ser indiretas ou obscuras; as fonéticas, aproximadas.
- Algumas pistas semânticas tornaram-se arcaicas: muitos caracteres usam o componente 貝 “concha” para o campo semântico “riqueza”, e 竹 “bambu” para “documentos”, porque na China antiga búzios eram moeda de troca e tiras de bambu, material de escrita.
- Alguns componentes diferentes foram simplificados em formas visualmente idênticas (como “carne” 肉 e “barco” 船, que em posição presa assumem a forma 月,

<sup>44</sup> O histórico debate entre os sinologistas Boodberg e Creel se deu, em grande parte, devido a esta indeterminação: onde Creel via claras associações semânticas, Boodberg via “fantasias surrealistas e farsescas” (LURIE, 2006).

idêntica ao componente “lua”). Em alguns casos, pictogramas arcaicos foram simplificados e agora têm a mesma forma que algum componente moderno, sem nenhuma relação semântica ou fonética. Na prática, isso significa que as relações entre componente e som ou sentido sempre podem ser polivalentes, de um para muitos (como 句 = {gǒu, jù} visto acima).

- Cada componente pode agir como pista fonética, semântica, ambas, ou nenhuma. A priori, não há como determinar mecanicamente o papel de um componente. Em geral, os semânticos ficam à esquerda ou em cima, e os fonéticos ou fonosemânticos, à direita ou embaixo; mas existem exceções.
- Mudanças históricas nos sons, sentidos, e formas gráficas erodiram associações.
- Caracteres que caíram em desuso criaram componentes “órfãos”: Se um componente originalmente fonético permanece em uso em apenas um caractere, ele não age mais como pista mnemônica, pois precisa ser decorado individualmente.

### 3.3.3 Exceções e desvios do princípio morfográfico

Vimos na seção 3.2 (p. 56) que a escrita portuguesa é baseada no princípio fonográfico: ou seja, cada grafema (sinal unitário) representa um fonema (som unitário). Vimos também, na seção 3.2.2 (p. 59), que esse princípio não é absoluto; há casos que desviam do princípio, como a letra ⟨h⟩ ou os sinais matemáticos. Essas complexidades enriquecem a escrita, fazendo dela mais do que uma simples transcrição fonêmica.

Similarmente, a escrita chinesa também não se resume a uma transcrição dos morfemas. Sistemas como a pontuação ou notação matemática hoje são internacionais, e também empregados no chinês. Além disso, ela possui seus próprios tipos de desvio do princípio morfográfico. Veremos alguns nesta seção.

#### 3.3.3.1 Polivalência de leitura na escrita chinesa

Lembre-se que a expressão “polivalência de leitura” (3.2.2.3 à página 60) refere-se a grafemas que podem representar mais de um elemento, sendo que a escolha é determinada caso a caso; por exemplo, em português, a letra ⟨s⟩ pode representar o fonema /z/ (quando entre vogais) ou o fonema /s/ (nos demais casos).

Em uma escrita morfográfica, polivalência de leitura significa que, dependendo do contexto, o mesmo grafema pode representar morfemas diferentes—em termos tradicionais, o caractere tem várias “leituras” em potencial. Via de regra, os caracteres chineses evitam polivalência; ou seja, no caso mais comum, sua leitura é sempre a

mesma.<sup>45</sup> Contudo, há um número não-trivial de caracteres com múltiplas leituras. Dos 12980 caracteres listados no dicionário eletrônico CC-CEDICT, 758 (5,84%) apresentam mais de uma leitura possível. Destes, 688 (90,77%) possuem exatamente 2 leituras (DENISOWSKI et al., 2016).

É preciso ponderar que estes números não estão balanceados por frequência; parece ser o caso que muitas dessas leituras raramente ocorrem.

Dentre os 758 caracteres com múltiplas leituras que encontramos, boa parte possui leituras semelhantes entre si; 274 diferem apenas no tom, e mais 144 apenas na consoante inicial, que é fonologicamente próxima. Tais pronúncias são linguisticamente interessantes, pois refletem um fenômeno do chinês clássico: um tipo de alomorfia, onde o som do morfema varia de acordo com seu papel sintático (BRANNER, 2003). Por exemplo, compare os verbos com os substantivos:

**Tabela 22 – Pronúncias polivalentes eruditas do chinês literário**

	<b>Pronúncia</b>	<b>Sentido</b> <b>Verbo</b>		<b>Pronúncia</b>	<b>Sentido</b> <b>Substantivo</b>
知	zhī	“saber”		zhì	“conhecimento”
處	chǔ	“morar”		chù	“local”
乘	chéng	“andar de carruagem”	乘	shèng	“carruagem”
陣	chén	“formar, arrumar”	陣	zhèn	“formação (militar)”

Esta distinção sintática de alomorfia é conservadora; preservada graças a uma longa tradição filológica, ela é pouco natural para os falantes atuais. De acordo com Branner (op. cit.),

Muitas leituras conservadoras tornaram-se totalmente estranhas para os ouvidos modernos [...] os dicionários mais novos de mandarim nem mesmo listam muitas dessas variantes tradicionais, e elas estão desaparecendo rapidamente do conhecimento popular.

De fato, dos exemplos que citamos na tabela 22, extraídos de Branner, as leituras alternativas de 知 e 陣 já não estão listadas no dicionário CEDICT.

Julgamos então válido considerar que a leitura polivalente, no sistema chinês, é excepcional; o sistema é concebido com leituras unívocas para cada caractere, e ele tende a assim se estabilizar. Mesmo dentre os caracteres com múltiplas leituras em potencial, em muitos casos essas leituras são alomorfes do mesmo morfema; ou seja, o

<sup>45</sup> Um dos motivos para isso é o processo gerativo dos compostos fono-semânticos (形聲 *xíngshēng* / *keisei*); este processo foi usado para eliminar polivalências, criando caracteres novos para resolver ambiguidades (cf. 3.3.2.4 à página 78).

caractere não é realmente polivalente no nível morfológico, mas sim representa regularmente um só morfema, independente de sua pronúncia (cp. o conceito de profundidade ortográfica, por exemplo no português *medic-*, discutido na página 63).

### 3.3.3.2 Plurimorfemia: um caractere representa uma sequência de morfemas

Alguns caracteres excepcionais representam uma sequência inteira de morfemas (MAIR, 2011; MAIR, 2015):

**Tabela 23 – Exemplos de plurimorfemia no chinês: um caractere denota múltiplos morfemas**

Palavra	Grafia morfológica regular	Grafia plurimorfêmica
1 <i>qiān-wǎ</i> mil-watts “quilowatt”	千瓦	仟
2 <i>qiān-kè</i> mil-gramas “quilograma”	千克	尅
3 <i>tú-shū-guǎn</i> ilustrações-documentos-prédio “biblioteca”	圖書館	圉
4 <i>shè-huì-zhǔ-yì</i> social-união-guiar-sentido =sociedade-ideologia “socialismo”	社會主義	袂

Exemplos como 1 e 2 são sequências de dois caracteres comprimidos em um só, em uma espécie de abreviação gráfica. Já caracteres como 3 e 4 são composições originais, construídas a partir de alguns elementos gráficos que remetem à sequência original de caracteres.

Lembre-se de nossa definição de plurifonia (3.2.2.2 à página 59): no sistema de escrita português, a maior parte das letras representa um só fonema, mas uma letra como ⟨x⟩ pode representar uma sequência plural (/ks/). De forma semelhante, os caracteres chineses normalmente representam um só morfema, mas os exemplos acima representam uma sequência de vários deles. Por analogia, denominamos este fenômeno **plurimorfemia**.

Como cada morfema equivale a pelo menos uma sílaba, os caracteres plurimorfêmicos também são plurissilábicos. Isso significa que eles violam as duas partes da

equação 3.27 (p. 72); neste tipo de caractere, um só grafema representa vários morfemas, e também várias sílabas. Ao contrário da pluralidade que veremos no japonês, este fenômeno parece ser excepcional e periférico no sistema de escrita chinês. Segundo Mair, tais caracteres emergem espontaneamente no uso popular, mas tendem a ser vetados por dicionários e autoridades, em um esforço de padronização (MAIR, 2011; MAIR, 2015). Dos 12980 caracteres listados no dicionário eletrônico CC-CEDICT (DENISOWSKI et al., 2016), os únicos plurais são as contrações gráficas criadas para representar unidades métricas—todos atualmente em desuso:

**Tabela 24 – Listagem completa de caracteres chineses plurais no dicionário CC-CEDICT**

Palavra	Tradução	Grafia morfográfica regular	Grafia plurimorfêmica
<i>háo-kè</i>	mili-grama	毫克	𠄎
<i>háo-wǎ</i>	mili-watt	毫瓦	𠄎
<i>lí-kè</i>	centi-grama	厘克	𠄎
<i>lí-wǎ</i>	centi-watt	厘瓦	𠄎
<i>fēn-kè</i>	deci-grama	分克	𠄎
<i>fēn-wǎ</i>	deci-watt	分瓦	𠄎
<i>shí-kè</i>	deca-grama	十克	𠄎
<i>shí-wǎ</i>	deca-watt	十瓦	𠄎
<i>bǎi-kè</i>	hecta-grama	百克	𠄎
<i>bǎi-mǐ</i>	hectâmetro	百米	𠄎
<i>bǎi-wǎ</i>	hecta-watt	百瓦	𠄎
<i>qiān-kè</i>	quilo-grama	千克	𠄎
<i>qiān-wǎ</i>	quilo-watt	千瓦	𠄎

Vemos portanto que existem caracteres chineses plurais, que representam múltiplos morfemas e múltiplas sílabas; porém, eles são exceções. Na grande maioria dos casos, cada caractere representa um só morfema e um só sílaba (eq. 3.27, p. 72).

### 3.3.3.3 O mito do monossilabismo? Caracteres sub-morfêmicos e as palavras-borboleta

Considere novamente a equação 3.27 (p. 72), reproduzida abaixo:

(3.27)  $1 \text{ hànzi} : 1 \text{ morfema} : 1 \text{ sílaba}$

Com isso queremos dizer que, no caso típico, cada morfema equivale a exata-

mente uma sílaba, e cada sílaba-morfema é representada por exatamente um caractere. Vimos na seção anterior o que acontece quando a relação “1 *hànzì* : 1 morfema” é violada—ou seja, quando um caractere corresponde a mais de um morfema. E quanto à equivalência “1 morfema : 1 sílaba”? Ela parece bastante regular até agora, o que pode levar à conclusão que o chinês seja “monossilábico”: a cada sílaba um sentido. Essa impressão é reforçada pelo sistema de escrita, que sugere e sustenta uma cadência de leitura: a cada caractere uma sílaba, a cada caractere um sentido. Contudo, a classificação como “língua monossilábica” é disputada; a correspondência entre sílabas e morfemas nem sempre é exata, levando autores como Kennedy (1951), DeFrancis (1984, cap. 11) e Unger (2004) a argumentar contra o “mito do monossilabismo”.

O mito do monossilabismo apresenta-se como parte da crítica mais ampla contra a escrita chinesa: se nem toda sílaba é um morfema, então algumas sílabas não carregam sentido; porém, toda sílaba corresponde a um caractere; portanto, alguns caracteres não representariam sentidos, mas tão-somente sons. Por isso, DeFrancis (op. cit.) propõe que a escrita chinesa seria primariamente fonográfica—um mero “silabário desajeitado”.

Para entender esta crítica, será proveitoso nos debruçarmos por um momento sobre a questão do monossilabismo. Nesta seção, tentaremos mostrar que, por um lado, o chinês possui, sim, palavras polissilábicas; mas, por outro, que ele apresenta uma estrutura básica de um morfema por sílaba, e mesmo nos casos excepcionais onde sílabas não são morfemas, há uma tendência a convertê-las em morfemas, de forma que o sistema se estabilize na equação 3.27.

Começemos pela estrutura das palavras. O primeiro mal-entendido a ser desfeito é que *palavras* chinesas seriam monossilábicas; ou ainda que o chinês não teria morfemas presos, só palavras livres que podem ser recombinadas à vontade. Esta impressão, de fato um mito, é reforçada pela ortografia chinesa, porque ela torna difícil visualizar as palavras. Comparemos com o caso da escrita alfabética. Nos textos em português, as palavras são separadas por espaços, e por isso tornam-se unidades bastante claras para a comunidade de falantes letrados.<sup>46</sup> Já os morfemas não são marcados de nenhuma forma na escrita portuguesa, e portanto são unidades não óbvias; na seção 3.1.2 (p. 51) tivemos que usar vários testes para encontrá-los. Na escrita chinesa em *hànzì*, as demarcações funcionam precisamente ao contrário. Por um lado, cada caractere normalmente corresponde a um morfema, e portanto os morfemas tornam-se muito claros para a comunidade de falantes; de fato, no mandarim cotidiano, a palavra para “caractere”, *zì* 字, também é usada no sentido de “morfema” (PACKARD, 2000,

<sup>46</sup> Como vimos anteriormente, casos como ⟨em cima⟩ vs. ⟨embaixo⟩, ou ⟨por que⟩ vs. ⟨porque⟩, mostram que nem sempre a “palavra ortográfica” coincide com a “palavra linguística”. Contudo, a coincidência é comum o bastante para podermos dizer que a função básica dos espaços é separar as palavras.

p. 14–15). Mas, por outro lado, a ortografia não emprega espaços, nem nenhuma outra forma de demarcar palavras; portanto, não é óbvio onde as palavras começam e terminam.

De que maneiras podemos identificar as palavras chinesas? Observe os seguintes casos:<sup>47</sup>

(3.28) 学了  
*xué-le*  
 estudar-PRFV  
 “estudou, (quando tiver) estudado”

(3.29) 紅 鳥  
*hóng niǎo*  
 vermelho pássaro  
 “pássaro vermelho”

(3.30) 紅花  
*hóng-huā*  
 vermelho-flor  
 “açafraão (*Carthamus tinctorius*)”

Em 3.28, 学了 *xuéle*, a sílaba *-le* é um sufixo verbal que denota “ação completa” (aspecto perfectivo).<sup>48</sup> Trata-se de um morfema preso; ou seja, ele nunca aparece sozinho, mas apenas como sufixo acoplado a uma raiz ou base verbal.<sup>49</sup> Assim, está claro que *-le* não é uma palavra independente, e concluímos que *xuéle* é uma palavra formada por dois morfemas e duas sílabas.

Em 3.29, *hóng* e *niǎo* são palavras distintas. Sabemos disso porque ambas podem aparecer isoladamente; e, além disso, podemos trocar *hóng* “vermelho” por qualquer adjetivo, ou *niǎo* “pássaro” por qualquer substantivo, e o significado continuaria sendo uma composição simples dos dois, com o adjetivo qualificando o substantivo. Diferentemente disso, o exemplo 3.30, *hóng huā*, não indica uma flor vermelha qualquer, mas sim o nome de uma flor em particular, o açafraão. Não é possível prever esse sentido a partir dos morfemas constituintes, nem é possível trocar um dos morfemas sem alterar radicalmente o sentido. O sentido de *hóng huā* é opaco, ou seja, precisa ser aprendido

<sup>47</sup> Exemplos extraídos de Packard (2000).

<sup>48</sup> Um sentido equivalente ao *-nu* do japonês clássico, ou ao *-te shimau* do moderno.

<sup>49</sup> 了 *le* também aparece modificando constituintes sintáticos inteiros, maiores que uma só palavra, com o sentido de “mudança de estado” (她会走路了 *Tā huì zǒulù le* “**agora** ela consegue andar”). Estamos seguindo a análise morfológica de Packard, que considera o *-le* verbal, marcador de aspecto, como um afixo distinto e preso (p.74). O argumento básico é que o morfema tem sentidos distintos, um deles quando preso e outro quando livre; no sentido de aspecto perfectivo, ele aparece exclusivamente ligado a verbos. Ver a op.cit. para detalhes e outros morfemas semelhantes.

separadamente enquanto uma unidade própria. Parece então razoável tratar *hónghuà* como uma palavra, visto que o termo ocupa uma entrada própria no léxico mental.

Os exemplos 3.28–3.30 estão transcritos no formato das glosas morfológicas padrão Leipzig (cf. sec. 1.4 à página 38). Neste padrão, palavras distintas (como “pássaro vermelho”, *hóng niǎo*) são separadas por espaços, exatamente como no português e outras escritas alfabéticas modernas. O mesmo ocorre na ortografia alfabética do mandarim, o *pīnyīn*: os morfemas são agrupados em palavras (詞 *cí*), delimitadas por espaços. Porém, a escrita tradicional em caracteres *hànzì* não usa espaços, e isso cria um problema na hora de transcrevê-la para o alfabeto *pīnyīn*: é necessário decidir onde começar e terminar cada palavras (ou seja, onde inserir espaços). Nos exemplos acima, vimos dois critérios para isso: i) se uma unidade pode ou não ocorrer sozinha (se é “presa” ou “livre”; cf. sec. 3.1.2, pp. 51 em diante) e ii) se o sentido composto é transparente ou opaco. Existem vários outros critérios possíveis, envolvendo fenômenos que ocorrem em nível de palavra quanto à fonologia (tons, pausas), morfologia, sintaxe etc.; uma discussão completa escaparia ao escopo desta dissertação (para isso ver Packard (2000)). O ponto importante para nós é que o chinês não pode ser descrito como “monossilábico” no nível lexical; por quaisquer critérios que escolhamos, o vocabulário da língua claramente inclui inúmeras palavras polissilábicas—de fato, o tamanho mais comum das palavras é 2 sílabas (KENNEDY, 1951; DEFRANCIS, 1984).

Concluimos, portanto, que o monossilabismo do chinês é um mito—no que diz respeito às palavras. Porém, e no nível dos morfemas; seria razoável descrevê-los como monossilábicos? Tudo o que vimos até agora indica ser esta a regra geral; em seu estudo contemporâneo sobre a morfologia do chinês, Packard (op.cit., pp. 129–130, grifos nossos) conclui que

Os componentes das palavras chinesas são **quase invariavelmente** morfemas distintos que consistem de sílabas únicas—o morfema e a sílaba são **quase que** coextensivos em mandarim.

Consideremos as exceções ao “quase que”. Compare as seguintes palavras:

(3.31) 睡眠  
*shuì-mián*  
 dormir-dormir  
 “dormir”

(3.32) 睡衣  
*shuì-yí*  
 dormir-roupa  
 “pijama”

(3.33) 失眠  
*shī-mián*  
 perder-dormir  
 “insônia”

(3.34) 蝴蝶  
*húdié*  
 “borboleta”

O verbo chinês “dormir”, *shùimián* (3.31), é composto por dois morfemas (semi-) sinônimos: 睡 *shuì*- “dormir, deitar-se” e 眠 *-mián* “dormir, hibernar”. A existência desses dois morfemas é comprovada por sua presença em outras palavras, como as citadas *shuìyí* “roupa de dormir” = “pijama”, e *shīmián* “deixar de dormir, perder o sono” = “insônia” (muitos outros exemplos poderiam ser listados).

Palavras como *shùimián* são compostas a partir de dois sinônimos (ou semi-sinônimos): “dormir + dormir = dormir”. Tais **compostos sinonímicos** são incomuns em línguas europeias como o português, mas muito frequentes no léxico chinês (PACKARD, 2000, pp. 88–89,93). A ortografia dos compostos sinonímicos é totalmente regular e morfológica: cada caractere representa um morfema e uma sílaba.

Contudo, considere o exemplo 3.34, *húdié* “borboleta”. À primeira vista, ele pode parecer outro composto sinonímico: 蝴 *hú*- “borboleta” + 蝶 *-dié* “borboleta” = 蝴蝶 *húdié* “borboleta”. Somos levados a pensar isso não só por esse tipo de palavra ser comum na língua, mas também pela grafia, paralela à de palavras como *shùimián*. De fato, se consultarmos um dicionário tradicional de caracteres como o *Kangxi*, encontraremos entradas independentes para 蝴 *hú* e 蝶 *dié*, ambas com o sentido de “borboleta”. Deduziríamos então que *hú* e *dié* são dois morfemas distintos, sinônimos de “borboleta”.

Há porém um problema: nos textos clássicos, os caracteres 蝴蝶 sempre ocorrem juntos, nesta ordem. Ou seja, a palavra para “borboleta” sempre é *húdié*, e nunca os caracteres individuais aparecem em outras palavras. Portanto, falta evidência linguística para considerar *hú*- ou *-dié* morfemas independentes; ao invés disso, *húdié* inteira parece ser um morfema dissilábico, significando “borboleta”.

Tal foi o influente argumento de Kennedy (1951, 1964), para recusar o monossilabismo dos morfemas. Kennedy culpa os caracteres chineses pelo surgimento desse “mito”. Adaptando um exemplo do autor para o português (1951, p. 165),

[...] se nossa escrita fosse separada por sílabas, como a chinesa, ela causaria uma impressão mais ou menos assim: (bei jar te no jar dim a le gre men te). Neste caso, nós também teríamos menos consciência das delimitações das palavras. Mais agravante ainda seria se cada sílaba não fosse sintetizada a partir de unidades sonoras menores [letras representando fonemas], mas sim, como no chinês, representada em sua

totalidade por um símbolo arbitrário. Disso se segue que os sinais representando *jar* em “beijar” e *jar* em “jardim” não necessariamente seriam semelhantes visualmente—algo que, em si, seria suficiente para dar a sensação de que sejam entidades distintas, ao invés de meramente maneiras diferentes de escrever a mesma sílaba. E mesmo que, no começo, o mesmo símbolo fosse usado para ambas, o gosto dos calígrafos chineses sempre foi tal que os dois sinais silábicos para ⟨bei jar⟩ logo seriam ornamentados por lábios e coraçõezinhos, enquanto que os de ⟨jar dim⟩ seriam decorados com roseiras e arabescos florais. Esses elementos adicionais são conhecidos como “determinativos” ou “radicais”, e aparecem na grande maioria dos caracteres, indicando campos semânticos gerais, tais como “palavras que têm a ver com água”, ou “comida”, ou “árvores”. Através de tais sinais, inventa-se o “*jar* de *beijar*; *jar*+boca” como entidade distinta do “*jar* de *jardim*; *jar*+plantas”. Cada sílaba adquire não só uma personalidade própria, mas também o poder de sugerir a palavra inteira da qual foi extraída. Porém “*jar* de boca” é um caractere que só aparece na palavra “beijar”, e “*jar* de plantas” somente em “jardim”. E quando um imperador ordena que se cataloguem todos os sinais gráficos, os dicionaristas criam quatro entradas distintas, da seguinte forma:

- “*Bei*+boca”: usado em “beijar”.
- “*Jar*+boca”: usado em “beijar”.
- “*Jar*+planta”: usado em “jardim”.
- “*Dim*+planta”: usado em “jardim”.

Se as palavras estiverem vivas na fala, este formato não causará nenhum problema, e servirá ao propósito de ilustrar qual é a ortografia correta. O fato que o caractere “*jar*+planta” é sozinho suficiente para representar a palavra “jardim” pode levar até a um tipo de abreviação literária; mas essa abreviação, por si só, não faz de *jar* uma palavra real, isto é, parte da fala. Porém, assim que uma palavra dissilábica torna-se obsoleta, seu destino está selado (e é um ponto crucialmente importante que a literatura chinesa foi quase toda escrita em uma língua já obsoleta). Imagine um momento em que a palavra “jardim” tivesse desaparecido por completo da fala cotidiana. Um ex-bandoleiro tornado poeta talvez fosse folhear um dicionário, e encontrasse ali o caractere *jar* repleto de flores e rosinhas, e talvez a essa altura já com uma cerquilha em volta. Quem o culparia por concluir que este lindo símbolo representaria uma antiga palavra *jar*, cujo sentido seria “cercadinho onde se plantam flores”? Poder-se-ia levantar a objeção que o poeta reconsideraria sua conclusão ao ler textos antigos, da época que a palavra “jardim” estava viva, e encontrasse neles a palavra completa, por extenso. Mas isso, ao contrário, só daria mais corda à sua imaginação. Por ter partido do pressuposto que *jar* e *dim* seriam palavras independentes, ele logo concluiria que “*jar*+planta”, por associação com “arranjar” e “alojar”, refere-se à uma área de cultivo deliberado, enquanto que “*dim* de plantas” refere-se especificamente a uma plantação *diminuta*. Nenhum sistema de escrita se presta tão bem à falsa etimologia, e tal “vantagem” foi amplamente explorada pelos chineses. Por exemplo, um verso em um poema muito antigo do *Livro das Odes* poderia ser parafraseada como “cri-cri cantam os grilos”. Comentaristas posteriores nos explicam gravemente que “grilos” [蟋蟀 *xīshuài*] comem plantas, sendo que os *gris* [蟋 *xī*] comem os talos e os *los* [蟀 *shuài*], as folhas. Ou então considere uma ave mitológica, a “fênix” [鳳凰 *fènghuāng*]. Segundo os dicionários o *fê* [鳳 *fèng*] é o macho e a *nix* [凰 *huāng*], a fêmea. Espécimes de tal tipo de “análise” poderiam ser listados às centenas.

Ou seja, o argumento de Kennedy é que, em palavras como 蝴蝶 *húdié* “borboleta” (e também 蟋蟀 *xīshuài* “grilo”, 鳳凰 *fènghuāng* “fênix” etc.), cada caractere não representa um morfema, mas sim uma mera sílaba sem sentido próprio; isso é provado sobretudo de forma diacrônica, pelo fato que nos textos clássicos os caracteres só aparecem como parte de uma única palavra polissilábica. DeFrancis enfatiza este ponto ao categorizar os caracteres chineses como “morfossilábicos”, ao invés de “morfográficos”: eles não representam apenas morfemas, mas também sílabas; aliás, para o autor, a representação de sílabas é que é primária, não a de morfemas (DEFRANCIS, 1984, pp. 124–125).

É interessante notar que os dois caracteres de 蝴蝶 *húdié* “borboleta” são ambos caracteres do tipo composto fono-semântico (cf. p. 83), com um componente para o som e outro para o sentido; que, além disso, ambos contém o mesmo componente semântico, 虫 “inseto”; e que diferem apenas nos componentes fonéticos, cada um indicando a pronúncia distinta de cada sílaba. Ou seja, eles têm uma estrutura da forma SA-SB, onde S representa o sentido, e A-B os sons. Esta estrutura gráfica parece comum também para outras “palavras-borboleta” (monomorfêmicas e dissilábicas, como *húdié*):

**Tabela 25 – Grafia SA-SB de palavras-borboleta: morfemas dissilábicos cujos caracteres partilham o mesmo componente semântico**

Palavra	Grafia	Componentes		
		Semântico	Fonético A	Fonético B
<i>húdié</i> “borboleta”	蝴蝶	虫 “inseto”	胡 <i>hú</i>	葉 <i>yè</i>
<i>xīshuài</i> “grilo”	蟋蟀	虫 “inseto”	悉 <i>xī</i>	率 <i>suài</i>
<i>shānhú</i> “coral”	珊瑚	玉 “jóia”	刪 <i>shān</i>	胡 <i>hú</i>
<i>pútao</i> “uva”	葡萄	艹 “vegetal”	匍 <i>pú</i>	萄 <i>táo</i>

Porém, nem toda palavra escrita na forma SA-SB é uma palavra-borboleta. 睡眠 *shuìmián* “dormir”, analisada no começo desta seção, possui estrutura gráfica SA-SB; mas cada sílaba é inegavelmente um morfema distinto.

A maioria dos morfemas dissilábicos parece se enquadrar em duas categorias (HANDEL, 2015, p. 139). A primeira são empréstimos de origem estrangeira, como 巴西 *Bāxī* “Brasil”, ou 涅槃 *nièpán* “nirvana” (no budismo). Alguns destes já estão naturalizados há séculos ou milênios, como *pútao* “uva” < elamita *budāwa*. A segunda categoria são palavras de simbolismo sonoro,<sup>50</sup> com repetição onomatopéica, como em 蟋蟀 *xīshuài* “grilo” < chinês antigo \**srit-srut*, ou 蝙蝠 *biānfú* “morcego” < \**pen-puk* (SCHUESLER, 2007, p. 165,524).

<sup>50</sup> Semelhantes aos *gitaigo* japoneses

De uma amostra aleatória de 200 caracteres, DeFrancis estimou que 11% representam sílabas não-morfêmicas (DEFRANCIS, 1984, p. 185). O autor argumenta que esta proporção é alta demais para ser considerada “exceção”, e portanto que o chinês não pode ser caracterizados como “monossilábico” em nível morfêmico. Há porém alguns problemas com esta análise. O primeiro é o tamanho da amostra. Realizamos uma contagem no dicionário CEDICT, na íntegra; de 13153 caracteres, 956 aparecem exclusivamente acompanhados de algum outro—ou seja, no máximo 7,27% caracteres são membros de palavras-borboleta.<sup>51</sup> Além disso, ambos os números são uma simples contagem de dicionário (contagem de tipos), sem levar em conta a frequência de cada caractere. Sabe-se, por exemplo, que a frequência de palavras monossilábicas é muito alta, o que altera as conclusões (FENG, 2009, p. 5). Trabalhando em nível lexical, Sproat (2000, 154–155) estima que o número total de palavras-borboleta não passa das 100.

Há ainda um problema mais profundo com o argumento das palavras-borboleta. Observe:

Tabela 26 – O morfema *dié*

Palavra	Sentido
鳳蝶科 <i>fèngdiékē</i>	<i>Papilionidæ</i> (família de borboletas)
蛺蝶 <i>jiádié</i>	<i>Nymphalidæ</i> (ibid.)
燕尾蝶 <i>yānwěidié</i>	Borboleta rabo-de-andorinha
蝶泳 <i>diéyǒng</i>	Nado-borboleta
蝶骨 <i>diégǔ</i>	Oso esfenoide (em formato de borboleta)

Notamos que, ao contrário do que argumentou Kennedy, 蝶 *dié* não é simplesmente uma abreviação literária: é um morfema vivo e produtivo, presente na língua moderna com o sentido de “borboleta”, e usado para criar palavras novas. É verdade que a primeira sílaba de *húdié*, 蝴 *hú*, não aparece em nenhuma outra palavra; mas, uma vez que *dié* é um morfema, um argumento mais simples seria que *hú* age como um morfema vazio, ou *cranberry morpheme*.<sup>52</sup> Por este argumento, caracteres como 蝴 *hú*

<sup>51</sup> Mesmo este número está superestimado. A contagem foi feita pelo seguinte método: buscamos caracteres que a) façam parte de uma palavra composta, e b) não apareçam em mais nenhuma palavra. Porém, isto também acaba contando caracteres livres, que podem aparecer como palavras independentes. O dicionário CEDICT não distingue morfemas presos de palavras livres; por isso, não foi possível diferenciar os dois casos. A proporção real de caracteres não-morfológicos será menor.

<sup>52</sup> Um morfema *cranberry* é um elemento que se combina com um morfema para formar uma palavra; porém, ele só aparece naquela palavra, e por isso não tem um sentido claro independente dela. Por exemplo, se o “galinheiro” é um local que abriga galinhas e “cupinzeiro” cupins, o que significa o “chiqu-” de “chiqueiro”?

O nome *cranberry* vem do exemplo canônico em inglês. O sufixo *-berry* nessa língua é usado para nomear frutos de baga, sendo adicionado a raízes como *blueberry* “baga azul” = mirtilo, ou *gooseberry* “baga-de-ganso” = groselha. Porém, em casos como *mulberry* (amora) ou *cranberry* (mirtilo-vermelho), os prefixos *mul-* e *cran-* são exclusivos a estas palavras, de forma que não têm significado próprio para o falante (exceto para um filólogo que tenha estudado sua etimologia). Na língua moderna, o morfema *cran-* em *cranberry* é um elemento vazio que serve apenas para diferenciar esta fruta de outras *-berries* como

não são fonogramas representando sílabas, mas sim morfogramas representando morfemas *cranberry*; afinal, o som *hú* só é representado pelo caractere 蝴 quando faz parte da palavra *húdié*. As sílabas individuais de *húdié* podem não ser originalmente morfemas, mas é válido argumentar que foram assim reanalisadas na língua atual (FENG, 2009, p. 9; HANDEL, 2015, p.141–142).

Uma maneira mais sofisticada de analisar o mesmo fenômeno, e que explica a tendência à reanálise morfológica, é postular que o morfema *húdié* possui duas formas alternativas, ou seja, dois alomorfes (cf. sec. 3.2.2.3.1, pp. 63–): em certos contextos ele se manifesta na “forma longa”, *húdié*, e em outros na “forma curta”, *dié*. Esta proposta é a das chamadas **palavras elásticas** (FENG, 2009; DUANMU, 2012; DONG, 2015). Observe:

	Forma longa	Forma curta	Sentido dos morfemas	Sentido da palavra
1	商店 <i>shāng-diàn</i>	<i>diàn</i>	comércio-loja	loja
2	老虎 <i>lǎo-hǔ</i>	<i>hǔ</i>	velho-tigre	tigre
3	煤炭 <i>méi-tàn</i>	<i>méi</i>	carvão mineral – carvão vegetal	carvão mineral
4	技术 <i>jì-shù</i>	<i>jì</i>	habilidade-técnica	habilidade
5	學習 <i>xué-xí</i>	<i>xué</i>	estudar-praticar	estudar

**Tabela 27 – Palavras elásticas no chinês (DONG, 2015, p. 1)**

Em todos os exemplos, a forma curta e a longa são ambas utilizadas como palavras independentes (morfemas livres), e são sinônimas. Considere a composição das formas longas. Na palavra 1, *shāngdiàn/diàn*, “(comércio-)loja” = “loja”, o prefixo *shāng-* é redundante: toda loja é comercial. Em 2, *lǎohǔ/hǔ* “(velho-)tigre” = “tigre”, o prefixo *lǎo-* é esvaziado de seu sentido normal; a palavra não quer dizer “tigre velho”, mas tão somente “tigre”, de qualquer idade, inclusive um filhote (de forma que “velho tigre” fica *lǎo lǎohǔ*). Em 3, também, o sentido “carvão vegetal” de *tàn* desaparece; *méitàn* não quer dizer “carvão mineral ou vegetal”, mas somente “carvão mineral”, exatamente como *méi*. 4 e 5 aproximam-se da noção clássica de compostos sinonímicos; mas a existência das formas monossilábicas demonstra que um dos morfemas é redundante.

É interessante notar que o uso das palavras dissilábicas está relacionado com fenômenos da sintaxe. Notoriamente, em uma composição de dois substantivos, não se usa a combinação curto-longo (“loja de carvão” pode ser *méitàn shāngdiàn*, *méitàn diàn*, ou *méi diàn*, mas não \**méi shāngdiàn*); ao passo que, em uma composição do tipo verbo-objeto, é a forma longo-curto que é preterida (DUANMU, 2012).

*blueberry*, *blackberry*, *mulberry* etc. Da mesma forma, pode-se argumentar que *hú* em *húdié* “borboleta” é um morfema vazio que serve apenas para formar a palavra *húdié* “borboleta”.

De onde vêm as palavras elásticas? Uma teoria muito difundida é a de Karlgren, segundo a qual no chinês antigo todas as palavras seriam monossilábicas (KARLGREN, 1923). Conforme o sistema fonológico se simplificou, muitos monossílabos tornaram-se homófonos, e para evitar essa homofonia excessiva, teriam sido acrescentadas sílabas extras. O problema com essa proposta é que ela não é suportada por evidência. As formas longas das palavras estão presentes nos textos antigos; o que acontece é que muitas vezes seu status enquanto palavras não havia sido reconhecido pelos filólogos clássicos (MAIR, 1992; MAIR, 1994). Dong (2015), em uma análise de corpus do chinês medieval, encontrou uma proporção de palavras elásticas equivalente à do moderno. Além disso, como Kennedy bem argumentou, mesmo no ápice histórico de complexidade da sílaba chinesa, seria difícil conceber o léxico inteiro de uma língua, que alcança centenas de milhares de palavras, representado apenas por monossílabos (KENNEDY, 1951). Parece claro que o chinês sempre teve palavras polissilábicas. Daí que, para Kennedy, como vimos na citação acima, as formas longas são as originais, e as formas curtas são uma espécie de criação artificial a partir da escrita, uma “abreviação literária”. Porém, esta segunda hipótese também não é validada pelos dados. Dong (op.cit.), usando métodos de corpus, calcula que pelo menos metade dos morfemas do mandarim moderno são elásticos, e Packard (2000) conclui que quase todas as suas sílabas carregam sentido; esses resultados vão de encontro à teoria das abreviações literárias, que “não fazem parte da fala”. Ao contrário, os alomorfes monossilábicos são a regra, não a exceção, e estão tão vivos na fala quanto as variantes longas.

Feng (2009) propõe que o chinês possui duas tendências aparentemente contraditórias: uma regra é que toda sílaba seja um morfema, e outra, que todo pé—unidade de ritmo na prosódia—seja dissilábico. A interação entre as duas regras faz com que morfemas livres monossilábicos ganhem uma sílaba extra, mesmo que seja um morfema vazio (*cranberry*), a fim de permitir seu uso em posições onde a prosódia exige um pé completo; ao passo que morfemas dissilábicos são reanalisados, atribuindo sentido às sílabas (como *húdié* “borboleta” → *dié* “borboleta” mencionado acima). Esta teoria é corroborada pelo extensivo estudo de Dong (op. cit.), que busca testar se as palavras elásticas advêm de tentativas de evitar homofonia (como propôs Karlgren), ou de requisitos prosódicos. Com análises históricas, dialetais e de corpora, todas as evidências mostram que a elasticidade não tem correlação com a homofonia, mas tem uma relação clara com a prosódia. Restrições sintáticas como as que mencionamos acima devem ser consequências do fato que certas posições sintáticas aceitam ou não ênfase prosódica. Para Link (2013), o chinês falado parece ser uma língua particularmente rítmica, ou pelo menos que tem um ritmo incomumente bem marcado, baseado no pé dissilábico; isso advém da interação entre morfemas monossilábicos e prosódia de ritmo binário.

Assim, nós concluímos que o chinês tem, sim, uma forte tendência ao monossilabismo, mas especificamente em nível de morfema, não em nível de palavras ou pés. Isso

torna a língua particularmente adequada para a representação por morfossilabogramas: como cada sílaba tem um sentido, é natural que cada caractere que corresponde a uma sílaba também seja interpretado como dotado de sentido.

#### 3.3.3.4 Caracteres chineses como fonogramas

Vimos que os caracteres chineses têm, em sua própria estrutura, componentes fonéticos—elementos gráficos que indicam a leitura, mas não o sentido. Ora, tais componentes são uma espécie de fonograma, como as letras do alfabeto. Como a maioria dos componentes também são caracteres livres, não é de se estranhar que tenha surgido a prática de usar *hànzì* para denotar o som, apenas, sem fazer referência ao sentido. Essa operação descarta o sentido e faz um pivô pelo som, a fim de alcançar um sentido novo; por isso, é parecida com o princípio do rebus (pp. 82–).

Um uso frequente de fonogramas é para grafar palavras estrangeiras. Por exemplo, a palavra sânscrita *Nirvāṇa* “extinção; despertar espiritual” foi representada em chinês como 涅槃. Normalmente esses caracteres são morfogramas de *niè* “enegrecer” e *pán* “prato”; mas aqui eles só denotam os sons *nièpán* = *Nirvāṇa*, deixando de lado os significados (a “iluminação espiritual” não significa “tingir pratos de preto”...). O mesmo vale para a grafia de Brasil > *Bāxī*, 巴西 (“apego” e “oeste”). Em alguns casos, a transcrição em fonogramas escolhe caracteres que sejam sugestivos: Coca-Cola é grafada 可口可樂 *Kěkǒu-Kělè*, que, se lido morfograficamente, seria “bom sabor, boa diversão”. Ao usar *hànzì* como fonogramas, o escritor têm inúmeros caracteres disponíveis para cada sílaba; é natural então escolher alguns cujo significado seja sugestivo, mesmo que a escrita não seja realmente morfográfica.

#### 3.3.4 O mito do ideograma

Considere a seguinte tradução de um “verso típico de poesia chinesa”,<sup>53</sup> proposta por Fenollosa et al. (1936) e transcrita por Ezra Pound, baseando-se nas anotações do autor (pp. 60–61):

<sup>53</sup> De autoria de Sugawara-no-Michizane (845–903).

Tabela 28 – Tradução de Fenollosa de um verso chinês

月	耀	如	晴	雪
LUA	RAIOS brilhante + penas voando	COMO mulher + boca	PURA sol + céu azul	NEVE chuva (nuvens ou pano sobre gotas d'água) + vassoura

Disco do sol com a ponta da lua [...] Acima à direita, figura abreviada de asas; abaixo, pássaro = voar. Fenollosa e Morrison ambos observam que é o pássaro de rabo curto.

Céu possivelmente contendo a ideia de tenda. O poeta evitou um “puro” contendo sol + vassoura. Movimento de varredura da neve; aparência da neve semelhante a uma vassoura.

Para Fenollosa, o caractere chinês não é um mero símbolo arbitrário, como as palavras da língua, mas se baseia em uma “vívida ilustração abreviada das operações da natureza”, e o texto chinês é uma sequência de símbolos visuais compostos, lidos “um após o outro, silenciosamente, pelo olho”. Por essa razão, o autor sequer fornece as leituras. Ao invés disso, ele acredita que o poeta está operando através de composições semânticas de componentes; “o poeta evitou o caractere ‘puro’ contendo ‘vassoura’” (provavelmente 淨), como se quisesse reforçar as ideias meteorológicas visualmente, por meio dos componentes.

Mas esta análise é bem diferente de como a poesia funciona para os leitores do chinês. O primeiro problema é que ela não é silenciosa; o testemunho dos leitores nativos é de que leem os caracteres como enunciados linguísticos, acompanhados da pronúncia, exatamente como os leitores de escritas fonográficas. Inúmeros experimentos demonstram que os centros fonológicos do cérebro são ativados durante a leitura de caracteres chineses; isso é verificável tanto pela mensuração direta de atividade neuronal, quanto por experimentos psicolinguísticos que demonstram efeitos fonéticos—por exemplo a pré-ativação (*priming*), ou erros de confusão de caracteres cuja pronúncia é semelhante (NAKAMURA et al., 2005; ERBAUGH, 2002; MATSUNAGA, 1994; KESS; MIYAMOTO, 1999).

Além disso, não é o caso que os caracteres representem cenas da natureza; as imagens descritas acima, como a varredura da neve, não passam pela cabeça do leitor;

o que vêm à mente é um enunciado linguístico específico, composto pelas palavras representadas. Assim, o poeta muito provavelmente não escolheu 晴 ao invés de 淨 porque queria trazer à mente uma ideia de “puro como o céu azul”, ao invés de “puro como algo limpo por uma vassoura”; os caracteres representam morfemas distintos, *sei* “dia ensolarado” vs. *jou* “puro, limpo”, e o poeta escolheu *sei* porque esta é a expressão que lhe veio à mente; \**sei-set*,<sup>54</sup> a “neve em um dia sol”, que é um par frequente ou *collocation* na poesia clássica.<sup>55</sup> O verso tem som, e era lido pelo autor como \**get-you jo sei-set*; e ele faz parte de uma estrofe que segue as regras bastante formais da métrica Tang, que inclui rimas e padrões tonais, de tal forma que cada linha inverte os tons que correspondem às respectivas sílabas da linha anterior.

Além de se equivocar sobre o que representam os caracteres chineses, Fenollosa também comete inúmeros equívocos na análise de seus componentes; por exemplo, em 雪, o elemento 冫 não representa “vassoura”, mas sim “mão direita”; 晴 não é “puro”, mas sim “clima ensolarado”, e o elemento à direita, 青, não é “céu”, mas sim “azul ou verde”. Mais importantemente, ele se equivoca quanto ao papel dos componentes—seu modo de funcionamento. Para Fenollosa, todos os componentes são semânticos. Como vimos na seção 3.3.2.4, a maioria dos caracteres (mais de 90% na contagem de tipos) incluem componentes com papel fonético. Acima, a palavra 如 *jo* “como” inclui o elemento 女 “mulher” não pelo sentido (pelo menos, não *somente* pelo sentido), mas porque este é pronunciado também *jo*, agindo como pista de leitura. Em 耀 “raio de luz”, o elemento à direita, 翟 “faisão” (não “voar”) talvez até sirva como pista semântica para a ideia da luz 光 que é como um pássaro; mas sem dúvida ele também tem o papel de sugerir a pronúncia *you*, também usado e.g. em 曜 *you* “dia da semana” ou 耀 *you* “brilho”.

Donde vemos que há dois problemas básicos com essa análise: que os caracteres chineses representariam ideias “puras”, abstratas, divorciadas da língua; e que seus componentes combinar-se-iam de forma exclusivamente semântica, sem base fonológica. Essas duas ideias estão relacionadas; combinadas, elas sustentam o equívoco de que a escrita chinesa seria

composta de símbolos e imagens, que, não tendo som algum, podem ser lidos em todas as línguas, formando uma espécie de pintura intelectual, uma álgebra metafísica ideal, que denota ideias por analogia, por relação, por convenção e assim por diante. (Jés Ko, 1776, apud DeFrancis (1984)).

<sup>54</sup> Como o poeta é japonês, estamos usando a pronúncia em japonês antigo do chinês clássico.

<sup>55</sup> Em defesa de Fenollosa, deve-se notar que, em alguns casos, a escolha de caracteres em poemas clássicos parece ter sido consciente do grafismo; e.g. em um poema do Man'yōshū, o autor justapôs vários caracteres repetindo o componente “folha”; em outro poema, há uma sequência semelhante de caracteres com o componente de “pássaro”. Mas esses casos parecem ser muito esporádicos; sem dúvida o objeto primário da poesia chinesa e japonesa são as palavras da língua, não os elementos gráficos.

Ou seja, o equívoco é, em uma palavra, que a escrita chinesa seria “ideográfica”, ao invés de morfográfica. Este “mito do ideograma” foi criticado e debatido sucessivamente por [Ponceau \(1838\)](#), [Boodberg \(1937\)](#), [Kennedy \(1958\)](#), [DeFrancis \(1984\)](#) et al. Uma retrospectiva dos debates é fornecido por [Lurie \(2006\)](#). O caso do japonês é discutido, dentre outros, por [Unger \(2004\)](#) e [Hannas \(1997\)](#); [Erbaugh \(2002\)](#) traz uma coletânea de autores de diversas áreas.

### 3.4 Fonografia vs. morfografia

A escrita chinesa, que à primeira vista parece ser puramente morfográfica, revela uma série de recursos fonográficos: componentes fonéticos, caracteres rebus, ou até o uso de *hànzì* como fonogramas. Similarmente, a escrita portuguesa, tida como uma das grafias alfabéticas mais foneticamente regulares, possui recursos que fazem referência ao sentido: heterografia, profundidade ortográfica, acento diferencial, pontuação semântica, algarismos arábicos... Concluímos que as duas escritas não são tão diferentes quanto parecem; ambas parecem incluir subcomponentes que referenciam tanto som quanto sentido. A diferença está na ênfase e na sistematicidade; a referência ao sentido na ortografia portuguesa é ocasional e irregular, enquanto que na chinesa ela é regra, e emprega mecanismos regulares (como os determinativos semânticos).

Vimos na seção 3.1 que a linguagem é composta por elementos unitários sem sentido próprio, chamados fonemas, que se re combinam em unidades maiores providas de sentido, os morfemas. Vimos também que a maioria das escritas baseia-se na representação dos fonemas (fonografia); enquanto que os caracteres chineses, bem como algumas outras escritas históricas, representam morfemas (morfografia). Como os morfemas são compostos por combinações de fonemas, o número de morfemas é muito maior. Contagens de fonemas dependem da análise, mas as estimativas mais generosas, alcançam, no máximo, cerca de 160 fonemas, no caso da língua taa (!xóõ) ([THE LANGUAGE ARCHIVE, 2015](#)). Números mais típicos incluem cerca de 44 fonemas no inglês britânico, ca. de 34 no português brasileiro, e ca. de 19 no japonês (28 se consideramos os sons longos como fonemas distintos) ([COMRIE, 1987](#)). É mais difícil precisar o número de morfemas de uma língua, pois nem sempre é trivial identificar computacionalmente os morfemas que compõe uma palavra; por exemplo, “anti-” denota um morfema em “anticorpo” mas não em “antigo”; a forma “manga” é compartilhada por dois morfemas distintos; “desfazer” é composto por “des-fazer”, mas “destacar” não implica “\*des-tacar”, enquanto “-onesto” em “desonesto” é um alógrafo do morfema “honesto”. Dificuldades à parte, parece claro que a contagem de morfemas ultrapassa a casa dos milhares.<sup>56</sup>

<sup>56</sup> Aplicando o analisador morfológico Polyglot ao léxico da base WordNet, estimamos pelo menos 5900 morfemas para o inglês (ver nosso programa `count-english-maybe-morphemes.py`, nos arquivos

Consequentemente, escritas morfográficas precisam de muito mais caracteres que as fonográficas. O alfabeto português usa 26 caracteres; mesmo contando as variantes maiúscula e minúscula, pontuação e letras com acento, o total não chega a 200—valores típicos para a maioria das línguas que, como o português, empregam o alfabeto romano.<sup>57</sup> Já na China, o leitor precisa conhecer no mínimo cerca de 3500 caracteres, com valores típicos girando em torno dos 6000; o número total de caracteres chega às dezenas de milhares, se consideramos o conjunto de todos os textos (cf. sec. 3.3.1 à página 73).

O grande número de caracteres implica que a morfografia tem uma carga maior na memória, no aprendizado, e mais complexidade no sistema de escrita. Todavia, ela permite recursos expressivos únicos, como os que veremos no capítulo 5.

## 3.5 Uma escrita mista: japonês

### 3.5.1 Introdução

Tendo estudado como funcionam a fonografia no português e a morfografia no chinês, estamos em condições de analisar o funcionamento da escrita japonesa. Partiremos do ponto em que ela é mais semelhante à chinesa, e seguiremos para suas particularidades.

Embora o foco deste trabalho seja o sistema atual (sincrônico), para entender sua estrutura, é produtivo acompanhar brevemente a história de sua formação. Exceto quando notado, a discussão que se segue é baseada em Seeley (1991).

### 3.5.2 O chinês no japonês: leituras *on*

Quando a escrita entrou no Japão pela Coreia, ela foi inicialmente usada para grafar a língua chinesa, apenas. O chinês clássico permaneceu em uso no Japão até o século XIX, como língua erudita escrita. Durante todo esse período, foi prática comum usar o chinês escrito em prefácios e comentários, documentos oficiais e *slogans*, e até na composição de poemas originais em língua chinesa (*kanshi*)—de fato, a primeira publicação poética do Japão foi uma antologia de poesia chinesa, o *Kaifūsō*, de 751 CE. Isso aconteceu a despeito do fato de que as línguas chinesas faladas nunca tiveram grande presença no Japão.<sup>58</sup> A situação na qual uma língua é usada com frequência na escrita, mas não na fala, é chamada **diglossia** (FERGUSON, 1966; SNOW, 2010). Em uma cultura

---

desta dissertação) (AL-RFOU, 2015; PRINCETON UNIVERSITY, 2010).

<sup>57</sup> Como é demonstrado, por exemplo, pelo padrão computacional ISO/IEC 8859-1:1998, que cobre a maioria das línguas européias com 191 caracteres. (ISO, 1998)

<sup>58</sup> Isso pode ser comparado, por exemplo, ao uso do latim clássico como língua escrita padrão na Europa, inclusive em países como a Irlanda medieval, que nunca fora terra de falantes românicos nativos.

diglósica, a língua de uso escrito é dotada de prestígio, tida por “clássica” ou mesmo “sagrada” (ROBERT, 2006), e serve como fonte de morfemas, empréstimos e calques à língua falada local, chamada **vernáculo**. Por outro lado, a língua escrita clássica não preserva sua fonologia própria, mas tem “fonologia parasítica”; ela é pronunciada com os fonemas do vernáculo. (O latim, por exemplo, foi pronunciado de uma maneira pela tradição britânica, de outra pelos francófonos, de outra pela Igreja, etc.)

Esse foi o caso do chinês clássico no Japão, onde ficou conhecido por *kanbun* (“texto chinês”).<sup>59</sup> A presença do *kanbun* resultou em uma importação massiva de morfemas e palavras chinesas; mais de 60% do léxico japonês é de origem chinesa (na contagem de tipos), e os morfemas chineses são usados para criar novas palavras, como acontece com as raízes gregas no português (SHIBATANI, 1990, p.142).<sup>60</sup> Essas palavras e morfemas entraram na língua falada não por contato linguístico falado direto, mas por via indireta, por meio da cultura escrita (FRELLESVIG, 2010, cap. 9, pp. 258–). Quanto à fonologia, ela foi inicialmente baseada em dialetos de prestígio da China, com força suficiente para induzir mudanças significativas na própria fonologia japonesa; mas, conforme os empréstimos se difundiram na fala, a fonologia original foi drasticamente reduzida e por fim adaptada ao vernáculo.<sup>61</sup>

A primeira consequência disso é que os textos em língua chinesa foram lidos, no Japão, com uma pronúncia adaptada à fonologia local. Esse tipo de leitura se chama *on*, ou “pelo som”. Por exemplo, a palavra 漢字 *hànzi*, “caractere chinês” é pronunciada *kanji* na leitura *on*. Nós usaremos o termo **sinítico** (“chinês, de origem chinesa”) para se referir aos morfemas assim chineses introduzidos desta maneira; às leituras *on* dos *kanji*, que representam estes morfemas; e às palavras com eles compostas, as chamadas “palavras chinesas” ou *kango* (termo que inclui tanto palavras oriundas da China propriamente, quanto aquelas criadas no Japão a partir das raízes chinesas). O estrato sinítico no japonês, considerado como um todo, é chamado **sino-japonês** (MIYAKE, 2003). Para ilustrar o efeito da adaptação fonética, considere novamente o verso em chinês citado por Fenollosa (página 101), de autoria do japonês Sugawara-no-Michizane (845–903):

<sup>59</sup> Mesmo na China, o chinês clássico escrito, chamado lá de *wényán* 文言, era uma língua distinta do chinês falado, caracterizando diglossia. Ver Mair (1994); Boiko (2013).

<sup>60</sup> Compare com o estrato de palavras românicas no inglês, estimado em 58% (FINKENSTAEDT; WOLFF, 1973). Note porém que em ambos os casos as palavras nativas são mais frequentes; a proporção média de ocorrências efetivas (*tokens*) de palavras de origem chinesa, em discurso falado japonês, é de cerca de 23,6%; e, na, escrita, 41,3% (MIYAJI; KAI, 2008, p. 17).

<sup>61</sup> Para detalhes sobre a introdução das pronúncias de prestígio da China no Japão, ver Miyake (2003).

Tabela 29 – Comparação de leituras chinesas e japonesas

	月	耀	如	晴	雪
Chinês médio	/ɲyat /	/ja /	/ná /	/ts <sup>h</sup> ieŋ /	/syat /
Japonês clássico	get /get /	you /jow /	jo /zo /	sei /sej /	set /set /
Japonês moderno	getsu /getsu /	yō /jo: /	jo /zo /	sei /sej /	setsu /setsu /
Mandarim moderno	yuè /jwe /	yào /jao /	rú /zu /	qìng /tɕ <sup>h</sup> iŋ /	xuě /ɕwe /
Cantonês moderno	jyut <sup>6</sup> /jy:t /	jiu <sup>6</sup> /ji:u /	jyu <sup>4</sup> /jy: /	ceng <sup>4</sup> /ts <sup>h</sup> ɛ:ŋ /	syut <sup>3</sup> /sy:t /

A pronúncia do japonês moderno desenvolveu-se a partir do japonês clássico, que por sua vez foi adaptado do chinês médio. Acrescentamos o mandarim e o cantonês para permitir a comparação com outros desenvolvimentos. (Note que os padrões para notação em alfabeto romano escondem algumas das relações, de forma que é melhor comparar as pronúncias pela notação IPA (entre barras). A letra ⟨j⟩ acima, na notação do chinês médio e cantonês, representa o mesmo som que ⟨y⟩ do japonês (/j /); o ⟨r⟩ do mandarim denota um som próximo ao ⟨j⟩ do japonês, etc.).

### 3.5.2.1 Estratos sinóticos: *go, kan, tōsō*

Os morfemas sinóticos muitas vezes foram reintroduzidas em épocas diferentes, ou sofreram transformações fonológicas que criaram versões alternativas, de forma que vários estratos sinóticos hoje coexistem na língua japonesa. Isso é comparável, por exemplo, à coexistência no português de “távola” e “tábua”, “plano” e “chão”, “pluvial” e “chuva” (cada par oriundo de uma mesma palavra latina—*tabula, plānum, pluvia*—mas aportuguesado em diferentes estágios de mudança fonética). Porém, há uma diferença importante do português: no japonês, devido à morfografia baseada no chinês, os vários estratos terminaram grafados de forma idêntica (são homógrafos). Seria como se as palavras “távola” e “tábua” fossem ambas grafadas identicamente como *tabula*, e a pronúncia fosse adaptada de acordo com o contexto.

No japonês, isso resultou em um nível muito maior de polivalência de leitura que o chinês—isso é, cada caractere rotineiramente pode dispor de múltiplas leituras em potencial. Por exemplo, considere as três pronúncias sino-japonesas (*on*) do caractere 行, “ir, linha”:

Tabela 30 – Exemplos de múltiplas leituras sinóticas atribuídas a um único caractere

Palavra	Pronúncia	
行列	<i>gyō-retsu</i>	ir-linha = “fazer fila”
行進	<i>kō-shin</i>	ir-avançar = “marchar”
行脚	<i>an-gya</i>	ir-perna = “peregrinar”

Nas três palavras acima, o mesmo caractere 行 é pronunciado variamente como *gyō* (<*gyaw* < \*/*gan*/), *kō* (<*kaw*) e *an* (< *æn*<sup>H</sup>; cf. coreano /ha:ŋ, hɛŋ/, cantonês /haŋ/ etc.). Essas pronúncias são de dialetos e estágios diferentes do chinês antigo, e na tradição japonesa classificam-se como da “pronúncia da Dinastia Wu” (*gō'on*), “da Dinastia Han” (*kan'on*), e “das Dinastias Tang-Song” (*tōsō'on*). A classificação tradicional é convencional, não historicamente precisa. É possível a um caractere carregar mais de uma leitura de cada categoria. Miyake (2003) descreve em detalhes como cada sistema de pronúncia surgiu, atribuindo-os principalmente a:

**Tabela 31 – Principais fontes dos estratos de leituras *on***

<b>Estrato <i>on</i></b>	<b>Fonte</b>
<i>Go'on</i>	Chinês antigo tardio e chinês médio-baixo ( <i>Late Old Chinese, Early Middle Chinese</i> ), via adaptação coreana (sino-Paekche).
<i>Kan'on</i>	Alto chinês médio ( <i>Late Middle Chinese</i> ) no dialeto de Chang'an.
<i>Tōsō'on</i>	Alto chinês médio das dinastias Yuan, Song e posteriores.

Do ponto de vista sincrônico, a consequência importante para nós é que cada caractere pode ter várias leituras; e que, diferentemente do chinês, esta polivalência tornou-se sistemática.

Do ponto de vista linguístico, *gyō*, *kō* e *an* são morfemas distintos, ainda que de mesma origem (cognatos) e sentido (sinônimos). O fato de que eles são, na escrita japonesa, escritos da mesma forma, acaba tendo consequências para a organização do léxico mental; os vários morfemas agrupados sob o mesmo caractere também são agrupados mentalmente em um mesmo conjunto, de forma que, por exemplo, ler um deles pré-ativa (*prime*) os demais (JOYCE, 2004). Do ponto de vista do falante, o *kanji* torna-se uma espécie de símbolo visual ou ponto focal desses morfemas sinônimos.

Note que a variação acima é determinada pelo léxico. A palavra *gyōretsu* sempre é *gyōretsu* e nunca *\*kōretsu* ou *\*anretsu*, ainda que essas seja leituras potenciais do mesmo caractere. Ou seja, durante o letramento, o aprendiz de japonês precisa internalizar a ortografia específica de cada palavra (semelhantemente à forma com que o aprendiz brasileiro precisa aprender qual palavra usa e.g. ⟨s⟩ ou ⟨z⟩ para grafar o som /z/). Não é suficiente meramente decorar as leituras dos *kanji* fora de contexto.

### 3.5.3 Fonografia com o *kanji*

Quando a escrita chinesa foi introduzida no Japão, ainda em língua chinesa, um problema que logo surgiu foi a representação dos nomes próprios japoneses. A solução adotada foi a mesma que os chineses empregavam para representar nomes e palavras estrangeiras em geral (por exemplo, do sânscrito): usar os caracteres como fonogramas,

descartando seu sentido (cf. sec. 3.3.3.4).<sup>62</sup> Por exemplo, um dos primeiros topônimos dentre os que viriam a designar o Japão foi “Yamatai”, que foi grafado em *kanji* como 邪馬台. Lidos morfograficamente, os caracteres significariam “maligno”, “cavalo” e “pedestal”, o que pela sintaxe resultaria em “pedestal de cavalos malignos”—o que certamente não parece um nome auspicioso! Acontece que esses caracteres, já na China antiga, eram com frequência usados como fonogramas, para grafar as sílabas *ya*, *ma* e *tai*. Há evidência filológica que os fonogramas empregados no Japão descendem daqueles empregados na Coreia e na China (BENTLEY, 2001).

### 3.5.3.1 Terminologia da fonografia em *kanji*: *ateji*, *kana*, *Man’yōgana*

A fonografia baseada em *kanji* recebeu diversos nomes em japonês. No período histórico, ela era chamada de *kana* (“nomes emprestados”), em oposição a *mana* (“nomes verdadeiros”) (LAMARRE, 2000). Posteriormente, a palavra *kana* passou a ser usada para designar uma categoria distinta de caracteres simplificados, especializados para este uso (cf. sec. 3.5.5). Hoje em dia o termo mais comum para fonografia em *kanji* é *ateji*, “caracteres atribuídos” (BRIGHT; DANIELS, 1996, p. 210).<sup>63</sup> No sistema moderno, esse tipo de fonografia *ateji* limita-se a alguns termos consagrados, como *mechakucha* “confusão”, grafada como 滅茶苦茶. Lidos como morfogramas, estes caracteres denotariam os morfemas “extinção”, “chá”, “dor” e “chá”; mas aqui os sentidos são descartados, e cada fonograma representa apenas os sons de cada sílaba.

Retroativamente, o uso extensivo de fonogramas *kanji* na era clássica tem sido hoje denominado *Man’yōgana*—literalmente o *kana* do *Man’yōshū*, em referência à coletânea poética epônima do século XIII. Porém, o termo induz facilmente a mal-entendidos. Muitos parecem supor, a partir dele, que a fonografia teria sido inventada para escrever os poemas do *Man’yō*; mas, como vimos acima, ela na verdade existe desde os primeiros registros da escrita chinesa, antedatando em muitos séculos sua entrada no Japão.<sup>64</sup> Outros supõe que o *Man’yōgana* foi um primeiro trabalho escrito fonograficamente; mas a obra está longe de ser a primeira, e de toda forma, a maior parte de seu conteúdo foi escrita morfograficamente—pelo sentido, não pelo som (WAKISAKA, 1992).<sup>65</sup>

<sup>62</sup> De fato, no compêndio histórico chinês *Sòng Shū* 宋書 (“O Livro de Song”), composto em 492-493, nomes japoneses estão grafados em caracteres chineses fonográficos; como o nome da Imperatriz Himiko (<\*Pimikwo), grafado como 卑彌呼 \**pjie.mi.kwo*. Esta grafia, criada pelos chineses, não é atestada no Japão antigo (FRELLESVIG, 2010, p. 21).

<sup>63</sup> O termo *ateji* atualmente tem sido usado de forma mais genérica para indicar não apenas a fonografia, mas quaisquer usos não-padrão do *kanji*, tais como as técnicas expressivas discutidas no capítulo 5. Em linguagem técnica, a definição mais estrita acima permanece padrão.

<sup>64</sup> De fato, pode-se argumentar que os componentes fonéticos dos *kanji*, discutidos na seção 3.3.2.4, já constituem uma espécie de fonografia interna; a fonografia é intrínseca ao *kanji*, e os caracteres *kana* modernos podem ser vistos como simplesmente o componente fonográfico do *kanji* tomado isoladamente.

<sup>65</sup> A característica realmente particular ao *Man’yōshū* é precisamente o contrário do simples “*Man’yōgana*” fonográfico: a obra representa um ápice na grafia criativa, indireta, complexa e lúdica com *kanji*. Para alguns exemplos, ver Seeley (1991, pp. 49–54).

3.5.3.2 Fonografia em *kanji* e a representação do japonês

É fácil pensar que a prática da fonografia permitiria escrever não apenas nomes próprios, mas textos completos em língua japonesa: basta escolher um *kanji* para aproximar o som de cada sílaba japonesa, e ignorar seus sentidos. Tal técnica foi de fato empregada e serviu a esse propósito, mas era deveras insatisfatória para o registro de textos completos, devido ao grande número de caracteres necessários, cada um com grafia trabalhosa, de múltiplos traços. Esta dificuldade foi famosamente registrada no primeiro compêndio histórico-mitológico, o *Kojiki* (ca. 711–712) (OGIHARA; KŌNOSU, 1973). O autor, Ō-no-Yasumaro, nota no Prefácio (um ensaio escrito em chinês, em elegante “prosa paralela” clássica) que (op. cit., pp. 47–49; LURIE, 2001, pp. 301–303):

然、  
 上古之時、言意並朴；  
 敷文構句、於字即難。  
 已因訓述者、詞不逮心；  
 全以音連者、事趣更長。  
 是以今、  
 或一句之中、交用音訓；  
 或一事之內、全以訓錄。  
 即、  
 辭理叵見、以注明；  
 意况易解、更非注。

Todavia,  
 Na Alta Antiguidade as palavras e sentidos eram claros;  
 esparramando orações e construindo frases, capturá-los na escrita é difícil.  
 Se explicados pelo método do sentido [*kun*/morfografia], o sentido não alcança o âmago;  
 se vertidos pelo método do som [*on*/fonografia], a passagem termina muito longa.  
 Aqui,  
 eu por vezes combinei numa só frase os dois métodos;  
 eu por vezes grafei um trecho inteiro no método do sentido.  
 Assim,  
 quando a lógica das palavras era difícil de ver, clarifiquei-a com notas;  
 quando o sentido era fácil de discernir, de notas absteve-me.

O *Kojiki* registra vários poemas japoneses antigos, escritos pela técnica da fonografia. Dessa forma, o autor pôde preservar exatamente o som e a métrica, atributos essenciais para canções. Além dos poemas, o corpo dos relatos do *Kojiki* constitui o primeiro grande texto em língua japonesa. Yasumaro nota acima que, se ele empregasse fonografia pura para o livro todo, a grafia ficaria demasiado extensa, dificultando a leitura e compreensão. Por isso, ele combinou a fonografia com outra técnica, o *kun*—morfografia japonesa—descrita na seção 3.5.4 a seguir.

### 3.5.4 A tradução como parte integral da escrita: leituras *kun*

Vimos que as variações históricas e geográficas do chinês clássico resultaram na atribuição de vários morfemas a um mesmo *kanji* no Japão, criando uma relação de um-para-muitos: cada caractere pode possuir diversas leituras em potencial. As leituras que vimos até agora foram variantes de uma mesma palavra-fonte chinesa—aquela originalmente representada pelo caractere. Porém, a pluralidade de leitura no japonês não se limita a variações do termo original; palavras de origem japonesa também foram atribuídas aos mesmos *kanji*.

Vimos ainda que a primeira aplicação da escrita chinesa no Japão foi para representar a escrita chinesa; e que, quando surgiu o interesse em grafar palavras japonesas, o uso de fonografia pura foi considerado muito longo e cansativo. A alternativa foi adaptar a morfografia, e a maneira mais simples de fazê-lo surgiu no contexto de **glosas**—traduções anotadas às margens de textos chineses (SEELEY, op.cit., pp. 91–; LURIE, op.cit.). Para ilustrar como isso ocorreu, imagine um diálogo hipotético entre dois monges japoneses, um aprendiz de escriba e um instrutor:

- Como se lê este caractere chinês, 行?
- Se lê *kaw*. (> moderno *kō*)
- Certo... e, bem, pode me explicar o que significa?
- Significa *yuku* (“ir”).

O estudante anota então o sentido ao lado do caractere. A primeira forma de ler segue a pronúncia original, e por isso foi chamada de “leitura pelo som”, *on*. A segunda forma, que traduz o chinês para palavras nativas, foi chamada de “leitura explicativa”, *kun*. A lógica do processo *kun* toma a seguinte forma: 行 em chinês antigo se lia *kaw* (>*kō*); *kaw* significa “ir” = *yuku*; portanto, *yuku* se escreve 行.

Com o tempo, desenvolveu-se uma técnica formal de tradução em *kun*, a “leitura explicativa de texto chinês” (*kanbun kundoku*), que permitia converter quase que mecanicamente um texto chinês para o japonês, reordenando a sintaxe dos morfemas e vertendo-os para os equivalentes japoneses. O resultado era a “leitura” de um original chinês como se fosse em língua japonesa. Com a difusão deste método, o chinês clássico acabou sendo usado também como uma espécie de código indireto para representar o japonês (embora um texto em “chinês” assim composto carregue certas anomalias sintáticas que denunciam sua estranheiridade; cf. Aldridge (2011)).<sup>66</sup>

<sup>66</sup> Esta prática de leitura, que verte *impromptu* o texto da língua clássica escrita para a língua vernacular, possui paralelos interessantes com as chamadas *glosas emilianenses*, que se prestavam à leitura do latim como se fosse espanhol. Por exemplo, ambos os sistemas, *kundoku* e *emilianenses*, lançaram mão de sinais diacríticos para reordenar a sintaxe, e de marcas para acrescentar partículas gramaticais; e em ambos os sistemas, o que à primeira vista parece ser um texto em outra língua (latim, chinês) era na verdade lido fluentemente em vernáculo (espanhol, japonês).

Ainda hoje o *kun* permanece como a maneira japonesa típica de se engajar com textos em chinês clássico. Como ilustração, considere o seguinte rolo caligráfico:

**Figura 5 – Exemplo de leitura *kundoku*: “As flores não atrapalham o caminho”**



O autor encontrou este rolo em uma aula de cerimônia do chá, e não conseguiu lê-lo por desconhecer o terceiro caractere. Indagado a respeito, o professor (falante nativo de japonês, não-falante de chinês) “leu” ou “explicou” (*kun*) o texto da seguinte forma:

- (Apontando para #1) *Hana wa...* (“As flores” + TOP)
- (Apontando para #4) *michi wo...* (“o caminho” + OBJ)
- (Apontando para #3) *samatage-* (“atrapalhar”, raiz verbal)
- (Apontando para #2) *-zu* (NEG, sufixo negativo).

A frase original na verdade é um provérbio em chinês clássico, cuja sintaxe neste caso é idêntica à do português:

花 不 礙 路  
 huā bù ài lù  
 flores NEG atrapalham caminho

“As flores não atrapalham o caminho.”

Na “leitura explicativa” em japonês, o professor desempenhou, de forma semi-automática, uma série de adaptações:

- Converteu os itens lexicais sinóticos (que em sino-japonês, *on*, seriam pronunciados *ka*, *fu*, *gai*, *ro*) em traduções convencionais nativas, os *kun* (respectivamente: *hana*, *-zu*, *samatage-*, *michi*);
- Reordenou os itens da sintaxe chinesa, sujeito-verbo-objeto com preposições, para a ordem japonesa, sujeito-objeto-verbo com posposições/sufixos verbais;
- E acrescentou posposições de caso (*-wa*, tópico, e *-wo*, objeto direto), ausentes no original, aonde o japonês as esperaria.

Essas operações que o falante desempenhou de forma fluente e automática correspondem perfeitamente à prática histórica da leitura *kanbun kundoku*.

Crucial para o *kundoku* foi o estabelecimento de uma associação forte, convencionalizada, entre morfemas japoneses (como *hana* “flor” ou, no exemplo anterior, *yuku* “ir”) e caracteres chineses, junto com os morfemas sinóticos por eles representados (como 花 *ka* “flor” ou 行 *gyō/kō/an* “ir”). Uma vez que essas associações se cristalizaram, as leituras possíveis de cada *kanji* ficaram ainda mais polivalentes: um caractere como 行 agora poderia ser lido *gyō*, *kō*, *an* ou *yuku*. Com isso, passou a ser possível representar a língua japonesa usando *kanji* morfograficamente: os morfemas sinóticos são representados pelos caracteres originais, e os morfemas nativos, por caracteres que correspondem à tradução/glosa *kun* codificada pela tradição.<sup>67</sup>

Em termos linguísticos, uma “leitura *on*” é um morfema de origem sinótica, introduzido no japonês juntamente com o caractere que o representava na China. Já uma “leitura *kun*” é uma tradução de um morfema *on*, que por convenção foi associada ao mesmo caractere, e por conseguinte ao morfema sinótico.

Assim como um caractere pode ter várias leituras *on*, ele também pode ter várias leituras *kun*. O verbo 行 *kaw* do chinês antigo não significava apenas “ir”, mas também “realizar”, nuance que em japonês se traduz por *okonau*. Esta também passou a ser uma outra leitura *kun* do caractere 行. E mudanças fonéticas criaram ainda a forma moderna *iku* “ir”, que hoje coexiste com sua antecessora *yuku* “ir”.

<sup>67</sup> Para uma discussão da importância de tais práticas de “leitura explicativa” na formação da língua japonesa, ver Lurie (2011).

O resultado é um tipo de morfografia bastante diferente da chinesa: morfemas de mesma origem (como *yuku* e *iku*), de etimologia distinta (como *okonau* e *iku*), ou mesmo de línguas diferentes (como *kō* e *iku*) são todos homógrafos, denotados pelo mesmo caractere; e, por meio deste agrupamento simbólico, acabam também agrupados no léxico mental (JOYCE, 2004), sendo de forma ampla considerados como morfemas equivalentes ou de “mesmo sentido”.

Os empréstimos sinóticos são em sua maioria raízes usadas para compor palavras, enquanto que os morfemas livres tendem a ser nativos; graças a isso, emergiu uma regra geral que as leituras *on* tendem a ser usadas em posição presa (composições de *kanji*), enquanto que as *kun*, em posição livre ou de raiz única (caracteres *kanji* isolados). Isto é apenas uma tendência geral, com inúmeras exceções de ambos os lados.

#### 3.5.4.1 Tradução além da morfografia: plurimorfemia e os compostos *jukujikun*

Nem sempre uma tradução mantém o mesmo número de morfemas que a língua-fonte. Considere as seguintes palavras no chinês e no japonês:

**Tabela 32 – Leituras *kun* plurimorfêmicas**

	Chinês	<i>On</i>	<i>Kun</i>
雷	luài “trovão”	<i>rai</i> “trovão”	<i>kami-nari</i> deuses-voz = “voz dos deuses” = “trovão”
唇	zúin “lábio”	<i>shin</i> “lábio”	<i>kuchi-biru</i> boca-borda = “borda da boca” = “lábio”
政	tǔej “governo”	<i>sei</i> “governo”	<i>matsur-i-goto</i> governar-CON-COISA = “assuntos de governar” = “governo”

Observe que as palavras em chinês são monomorfêmicas, mas a tradução japonesa não. Quando essas traduções firmaram-se como leituras dos caracteres correspondentes, o resultado foi que esses caracteres passaram a poder representar uma sequência de dois morfemas seguidos. Trata-se do fenômeno da plurimorfemia, que vimos no contexto chinês à página 89. Porém, enquanto no chinês ele só ocorre muito excepcionalmente, no japonês ele ocorrerá sempre que a tradução precisar de mais morfemas que o original. Outros exemplos incluem 陷 *xiàn* “afundar, cair” > 陥る *ochi-iru* cair-entrar = “afundar, cair”; 覆 *fù* “revirar” > 覆る *kutsu-gaeru* sapatos-virar = “revirar”, etc.

O contrário também ocorre:

Tabela 33 – Leituras *kun* compostas (*jukujikun*)

	Chinês	On	Kun
大人	dâi-nin grande-pessoa “pessoa grande” =“adulto”	—	<i>otona</i> “adulto”
梅雨	muâi-yu ameixeira-chuva “chuva das ameixeiras” =“estação chuvosa”	<i>bai-u</i> ameixeira-chuva “chuva das ameixeiras” =“estação chuvosa”	<i>tsuyu</i> “estação chuvosa”

Aqui, uma palavra chinesa composta foi vertida para uma palavra japonesa monomorfêmica; contudo, a escrita original chinesa foi mantida. O resultado é que e.g. o morfema japonês *otona* “adulto”, que é indivisível, é escrito com dois caracteres, na versão morfológica de um dígrafo: 大人. Como o primeiro caractere normalmente representa o morfema de “grande”, e o segundo, o de “pessoa” (o que daria o sino-japonês \**daijin*), a grafia evoca o sentido de “pessoa grande”, ainda que isso não seja a análise da palavra japonesa para “adulto”.

Em japonês, compostos como esse são chamados *jukujikun*: leituras *kun* “maduras”, isto é, convencionalizadas. Em tais casos, não é possível dizer que cada caractere corresponde a uma parte da palavra; a leitura está atribuída a uma sequência de caracteres, tomada como um todo indivisível. Não é o caso que 大 tenha uma leitura \**oto-* e 人, \**-na*; mas sim que o par integral 大人 tem a leitura *otona*, que é um só morfema.

O uso do *jukujikun* estendeu-se para além das palavras compostas de origem chinesa. Por exemplo, a palavra portuguesa “tabaco” foi grafada como 煙草; caracteres que, em leitura nativa *kun*, normalmente representariam os morfemas *kemuri* “fumaça” e *kusa* “erva”. Mas, ao invés de denotar a sequência *kemurigusa*, ela indica a palavra *tabako*—que está representada pelo sentido combinado de *kemurigusa*, “erva de fumaça”, deixando de lado os fonemas desses morfemas. Da mesma maneira, a espada japonesa de bambu, *shinai*, é grafada como 竹刀; os caracteres regularmente seriam lidos morfograficamente como *take-gatana*, “espada de bambu”, mas aqui são lidos de uma vez como a palavra *shinai*. O princípio *jukujikun* é, portanto, que os caracteres sugerem o significado, mas não a pronúncia, que é arbitrária.

Como o *jukujikun* trabalha com palavras inteiras, ao invés de morfemas individuais, ele pode ser entendido como uma forma de **logografia**—a escrita pela palavra. E como os caracteres individuais contribuem para o sentido, mas não para o som, é um caso excepcional de “ideografia” ou **semantografia**, propriamente: escrita baseada no significado e não no significante.

### 3.5.5 Fonogramas japoneses: o *kana*

Vimos que Ō-no-Yasumaro reclamou que a fonografia pura é demasiado longa, pois exige escrever maior número de caracteres; por isso, ele limita-a às transcrições de poemas nativos, onde era essencial. Com o grande número de traços dos caracteres chineses,<sup>68</sup> grafar sílaba a sílaba era um trabalho manual custoso, e a leitura também tornava-se confusa. Em resposta a isso, sugeriram duas técnicas de grafia abreviada.

A primeira é a **cursivização**. A escrita chinesa já era dotada, há muito tempo, de variantes cursivas, que simplificam os gestos da escrita arredondando, conectando e abreviando traços. Como a fonografia precisa de mais caracteres, tais técnicas eram bastante propícias. Uma forma super-cursivizada, baseada sobretudo nos modelos do calígrafo Wang Xizhi, difundiu-se como notação fonográfica. Por exemplo, o caractere 己 “si mesmo”, usado como fonograma para *ko*, foi cursivizado como 𠄎; e 加, “adicionar”, fonograma *ka*, tornou-se 𠄎. Um nome clássico dessa notação super-cursiva fonográfica foi *onnade* “mão feminina”, pois era a única forma de escrita disponível às mulheres da corte—o ensino de *kanji* sendo-lhes, a princípio, vedado (mas note que os homens também usavam *onnade* rotineiramente, e algumas mulheres aprenderam a morfografia plena; famosamente, a autoria do *Relato de Genji*, Murasaki Shikibu, criticou a autora do *Livro do Travesseiro*, Sei Shōnagon, por ostentar seu conhecimento de caracteres chineses) (YODA, 2000). Inicialmente considerada algo informal, esta notação cursiva eventualmente tornou-se a forma padrão de representar palavras japonesas pelo som, sob o nome *hiragana* (“*kana* simples”).

Em paralelo a isso, os monges budistas, que trabalhavam escolasticamente com textos chineses (*kanbun*), também empregavam as formas cursivas, precursoras do atual *hiragana*. Além disso, eles desenvolveram uma notação para glosas fonéticas, que escreviam ao lado das colunas de caracteres chineses para facilitar o aprendizado e recitação (*kuntenbon*). Essa notação também foi baseada nos fonogramas já existentes; mas, ao invés de cursivizar, ela os abreviou extraíndo traços e componentes isolados, seguindo o modelo do *kugyōl* coreano (FRELLESVIG, 2010, p. 160). Por exemplo, 加 *ka* gerou 𠄎, e 己 *ko*, 𠄎 (tomando só a parte superior). Esta grafia recebeu o nome de *katakana* (“*kana* quebrado”).

Vimos anteriormente que o termo *kana*, “nomes emprestados”, originalmente não denotava um tipo de caractere, mas sim uma forma de usar os caracteres; i.e. o uso fonográfico de *kanji*. Mas o surgimento de caracteres específicos para *kana* (*hiragana* e *katakana*) terminou por criar um sistema de divisão do trabalho: ao contrário do chinês, que usa os mesmos caracteres tanto para morfografia quanto para fonografia, no japo-

<sup>68</sup> Os 13110 caracteres e componentes listados no dicionário KANJIDIC (BREEN et al., 2016) possuem uma média de 12.91 traços (mediana 13, moda 12, desvio-padrão 4.64; contagem de tipos, sem balancear por frequência).

nês moderno quase sempre a fonografia usa os caracteres *kana* próprios, visualmente distintos, e a morfografia, *kanji* não-abreviados. É importante notar, porém, que esse processo de especialização se deu lentamente; o surgimento dos dois tipos de *kana* não impediu que os fonogramas em *kanji* continuassem sendo usados por muitos séculos (FRELLESVIG, 2010, p. 160).

A existência de duas variantes paralelas, *hiragana* e *katakana*, permitiu mais subespecialização. Os papéis de cada um variaram no decorrer da história; atualmente, a fonografia japonesa usa o *hiragana* em situações normais e o *katakana* como uma notação alternativa, semelhantemente ao que aconteceu com as letras “romana” e “itálica” na Europa. Os usos mais comuns do *katakana* incluem estrangeirismos, onomatopeias, nomes de espécies biológicas, e ênfase discursiva.

As tabelas a seguir apresentam os *kana* atualmente em uso, com os fonogramas em *kanji* de onde se originaram (SEELEY, 1991, apêndice 3):

**Tabela 34 – Os *hiragana* atuais, sua pronúncia e *kanji* de origem**

	∅	K	S	T	N	H	M	Y	R	W	-
A	a あ 安	ka か 加	sa さ 左	ta た 太	na な 奈	ha は 波	ma ま 末	ya や 也	ra ら 良	wa わ 和	N ん 无
I	i い 以	ki き 幾	shi し 之	chi ち 知	ni に 仁	hi ひ 比	mi み 美		ri り 利	wi ゐ 為	
U	u う 宇	ku く 久	su す 寸	tsu つ 川	nu ぬ 奴	fu ふ 不	mu む 武	yu ゆ 由	ru る 留		
E	e え 衣	ke け 計	se せ 世	te て 天	ne ね 祢	he へ 部	me め 女		re れ 礼	we ゑ 恵	
O	o お 於	ko こ 己	so そ 曾	to と 止	no の 乃	ho ほ 保	mo も 毛	yo よ 与	ro ろ 呂	wo を 遠	

Tabela 35 – Os *katakana* atuais, sua pronúncia e *kanji* de origem

	∅	K	S	T	N	H	M	Y	R	W	-	
A	a	ka	sa	ta	na	ha	ma	ya	ra	wa	N	
	ア	カ	サ	タ	ナ	ハ	マ	ヤ	ラ	ワ	ン	
		阿	加	散	多	奈	八	万,未?	也	良	和	-
I	i	ki	shi	chi	ni	hi	mi		ri	wi		
	イ	キ	シ	チ	ニ	ヒ	ミ		リ	ヰ		
		伊	幾	之	千	二	比		利	井		
U	u	ku	su	tsu	nu	fu	mu	yu	ru			
	ウ	ク	ス	ツ	ヌ	フ	ム	ユ	ル			
		宇	久	須	?	奴	不	牟	由	流		
E	e	ke	se	te	ne	he	me		re	we		
	エ	ケ	セ	テ	ネ	ヘ	メ		レ	ヱ		
		江	介	世	天	祢	部		礼	惠		
O	o	ko	so	to	no	ho	mo	yo	ro	wo		
	オ	コ	ソ	ト	ノ	ホ	モ	ヨ	ロ	ヲ		
		於	己	曾	止	乃	保	毛	与	呂	乎	

- Vários *katakana* foram derivados de simplificações cursivas intermediárias (não ilustradas aqui).
- Alguns *katakana* foram tomados diretamente do fonograma *kanji*, sem abreviação gráfica: チ *chi*, ニ *ni*, ハ *ha*, ミ *mi*.
- O caractere ス *su* deriva de uma variante de 須 com o componente “três gotas d’água” 彡 à esquerda.
- ツ *tsu* já aparece abreviado cedo, e a derivação é incerta.
- O símbolo ン *N* é provavelmente original, criado especificamente para representar a nasal moraica, sem se basear em fonogramas *kanji*.

As duas tabelas acima estão dispostas em um esquema que alinha as consoantes por colunas e as vogais por linhas, em uma ordem tradicional chamada Tabela dos Cinquenta Sons (*Gojūonzu*); Note contudo que cada uma só tem 48 símbolos. O esquema *gojūonzu* deriva diretamente de tabelas fonológicas do sânscrito, nas quais os fonemas são agrupados segundo suas características distintivas (FRELLESVIG, 2010, p. 178).<sup>69</sup>

<sup>69</sup> Falamos da ordem *varṇamālā* “guirlanda das letras/fonemas”, empregada em escritas brânicas

Uma vez que as simplificações visuais do *kana* terminaram por se codificar, os textos japoneses gradativamente passaram a apresentar uma divisão de trabalho: *kanji* completos para a morfografia, e *kana* para a fonografia. Ou seja, o sistema completo passou a ter dois subsistemas, um seguindo o princípio fonográfico e outro o morfográfico. Tal fenômeno é denominado **digrafia**.

### 3.5.5.1 Irregularidades na fonografia do *kana*

O *kana* é uma fonografia bastante regular, seguindo quase que perfeitamente a fonologia da língua. Note que a romanização Hepburn, mais fonética do que fonológica, esconde algumas regularidades nas transcrições das tabelas acima: os sons que em Hepburn grafam-se *chi*, *tsu*, *shi*, *fu* são simplesmente os fonemas /ti, tu, si, hu/ (que são realizados como [tɕi, tɕu, ɕi, ɕu] na pronúncia padrão de Tóquio). Ou seja, os *kana* correspondentes são regulares.

Vejam as exceções. Os sons /wi, we, wo/ fundiram-se no japonês médio com /i, e, o/, mas permaneceram distintos na ortografia clássica (um caso de heterografia): *wi* ゐ, ヰ, *we* ゑ, ヱ e *wo* を, ヲ passaram a ser pronunciados da mesma forma que *i* い, イ, *e* え, エ e *o* お, オ, mas as palavras por muito tempo foram escritas etimologicamente (de forma, por exemplo, que “mulher” *onna* permaneceu sendo escrito をんな *wo.n.na* muito depois da fusão *o/wo*). Esses caracteres homófonos referentes ao antigo *w-* foram abolidos na reforma ortográfica do pós-Guerra, no século xx. Apenas o caractere *wo* を, ヲ permaneceu, e para uso exclusivo como marcador de objeto direto (pronunciado *o*, homófono de お, オ). Trata-se de um caso de heterografia que distingue sentido, baseada em etimologia, como *a* vs. *à* do português brasileiro.

Caso semelhante ocorre com os *kana* は *ha* e へ *he*. O fonema consonantal /h/ do japonês moderno é um reflexo de um antigo /p/, que sofreu abrandamento para /ɸ/ (o ‘f’ de dois lábios, ainda presente antes de /u/ em /hu/ = [ɸu]). Além disso, quando em posição medial (no meio de palavras), o gesto bilabial vocalizou-se para a semivogal /w/. O resultado foi que, por boa parte do período histórico, os *kana* da atual série *h-* eram lidos como *w-* quando mediais. A semivogal /w-/, por sua vez, desapareceu posteriormente, exceto quando antes de /a/; isso fez com que *hi, hu, he, ho* ひふへほ, ヒフヘホ adquirissem a pronúncia de vogal simples, *i, u, e, o*. A reforma ortográfica moderna abandonou tais pronúncias obsoletas; porém, は *ha* e へ *he* (apenas) mantiveram as pronúncias históricas *wa* e *e*, exclusivamente para denotar as partícula gramaticais de tópico e do caso lativo (direção de destino), respectivamente.

como o devanāgarī (CHAUDHURI, 1997). A ordem *gojūonzu* dos *kana* reproduz exatamente a sequência de vogais *a-i-u-e-o* do *varṇamālā*, e reproduz parcialmente a sequência geral das consoantes—palatal, dental, labial, líquidas—ainda que com algumas irregularidades (mas note que o /h/ moderno era um /p/, bilabial). O budismo esotérico japonês manteve vivo até hoje a escrita brâmica chamada *siddhaṇi* (japonês 梵字 *bonji*); mas ela nunca se difundiu no país fora desse contexto especializado, a despeito de sua comparativa simplicidade e regularidade fonológica.

3.5.5.2 *Kana*, sílaba e mora

Sabemos que os *hànzì* chineses representavam sílabas. As sílabas chinesas são mais complexas fonotaticamente que as japonesas, aquelas apresentando semivogais, ataques, codas e tons ausentes destas. Em parte devido ao contato com o chinês, e em parte devido a desenvolvimentos internos, o japonês do período Heian (século VIII) viu emergir a propriedade da **quantidade**, que distingue vogais e consoantes por sua duração (FRELLESVIG, 2010). Por exemplo, *oji* com /i/ curto é “tio”, enquanto *ojii*, com /i:/ longo, “avô”. Na linguística, a quantidade de uma sílaba é medida por um valor abstrato chamado **mora**: uma sílaba curta tem uma mora, e uma sílaba longa, duas. Em linguagem casual, muitas vezes se diz que a escrita *kana* seria um “silabário”; mas, propriamente falando, ela é moraica.<sup>70</sup> Considere a tabela abaixo:

Tabela 36 – Relação entre *kana*, sílabas e moras

Palavra	Sílabas	Moras	<i>Kana</i>
<i>Yokohama</i>	4: [jo.ko.ha.ma]	4	4: よこはま
<i>Hōō</i> “fênix”	2: [ho:.o:]	4	4: ほうおう
<i>Sensē</i> “professor”	2: [sē.se:]	4	4: せんせい
<i>Aoi</i> “azul”	2: [a.oj] <sup>71</sup>	3	3: あおい
<i>Atta</i> “havia”	2: [a.t:a]	3	3: あった

Nos exemplos acima, sempre que uma sílaba longa resulta em maior número de moras, a contagem de caracteres *kana* acompanha a quantidade de moras, não a de sílabas. Existem exceções à equivalência entre *kana* e mora, como as semivogais no ataque ou *glides*, que acrescentam um *kana* mas não contam moras (*kya* > ki.ya きゃ); e os processos de palatalização (*sha* > shi.ya しゃ). Essas exceções são marcadas pelo uso de *kana* em tamanho reduzido, ditos “descartados” (*sutegana*).

3.5.5.3 Diacríticos do *kana*: *dakuten* e *handakuten*

濁 *Daku*, literalmente “lama”, refere-se a um processo morfofonêmico que, no japonês moderno, resulta em vozeamento de consoantes. Por exemplo:

<sup>70</sup> Um erro relacionado consiste na difundida análise ocidental da poesia japonesa (haikai, *tanka*) como consistindo de versos de 5 e 7 sílabas. Na verdade, a métrica pede versos de 5 e 7 moras, mais curtos portanto que penta- e eneassílabos europeus. Na terminologia poética japonesa, as moras são chamadas de *on* “sons”, *haku* “batidas”, ou—refletindo a natureza moraica da escrita *kana*—*ji* “caracteres”.

<sup>71</sup> A sequência *oi* faz ditongo; cf. Kubozono (2015, cap. 5).

Tabela 37 – Processo de vozeamento *daku* e o sinal *dakuten*

<i>Kuni</i> くに	“país”	<i>Yuki-guni</i> ゆきぐに	“país-da-neve”
<i>Sakura</i> さくら	“cerejeira”	<i>Fuyuzakura</i> ふゆざくら	“cerejeira-de-inverno”
<i>Tana</i> たな	“estante”	<i>Hon-dana</i> ほんだな	“estante-de-livros”
<i>Hana</i> はな	“flor”	<i>Hi-bana</i> ひばな	flor-de-fogo = “faísca”

Vemos que, quando uma palavra entra no meio de uma composição, está sujeita a ter sua consoante vozeada: /g, z, d/ são pronunciadas com o mesmo gesto articulatorio que /k, s, t/ respectivamente, mudando apenas o momento em que a voz começa a vibrar. A única exceção é o /h/, que se converte em /b/. Isso se deve ao fato que o /h/ do japonês era originalmente um /p/, que vozeado torna-se /b/. O /b/ foi preservado devido ao efeito da “posição protegida”.

Note que a grafia *kana* usa um diacrítico, a “marca de *daku*” ou *dakuten* (゛), para representar o vozeamento (e, no caso dos *kana* da consoante /h/, a conversão em /b/). O diacrítico foi depois adaptado para representar a consoante estrangeira /v/, ausente do japonês, acrescentando-o aos *kana* da vogal /u/: う゛, ヴ. Em textos informais, este diacrítico pode ainda ser adicionado a vogais para representar uma enunciação forçada ou enfática, ou aos *kana* de /g/ para denotar o alofone nasal dessa consoante ([ŋ]).

No século xvi, os missionários jesuítas portugueses iniciaram operações no Japão, dedicando-se (como era praxe para a ordem) a estudar a língua e produzir gramáticas, dicionários e textos religiosos. Um problema que encontraram foi que, devido à mudança fonética parcial do /p/ para (então) /ɸ/, não havia maneira de distinguir quando os *kana* dessa consoante estavam representando de fato um /p/. Os missionários introduziram o “sinal de meio *daku*” ou *handakuten*—um pequeno círculo sobrescrito. Com isso, tornou-se possível grafar /pa, pi, pu, pe, po/ como ぱぴぷぺぽ. O sinal teve boa aceitação e continua em uso (FRELLESVIG, 2010).

#### 3.5.5.4 *Kana* e sons longos

As vogais longas /a:, i:, u:/ são escritas acrescentando-se um *kana* vocálico: /ka:, ni:, su:/ grafam-se ⟨ka.a, ni.i, su.u⟩ = かあ、にい、すう.

As vogais longas /e:, o:/ normalmente são escritas ⟨-ei, -ou⟩. Apenas em alguns

casos, por razões etimológicas, é que a ortografia grafa com o equivalente a *-ee* ou *-oo* (exemplos incluem /to:ru/ “passar” = to.o.ru = とおる ou /e:/ “sim, bom” = e.e ええ).

As consoantes longas (geminção) são escritas precedendo-as pelo caractere つ *tsu*—desde o pós-guerra, grafado em menor tamanho quando neste papel: *atta* あった, *zasshi* ざっし, etc. Esta prática provavelmente deriva do uso antigo de *tsu* < *tu* para representar a consoante final *-t* do chinês, que inicialmente manteve-se viva nos empréstimos (como no japonês medieval *it* “um”, atual *ichi*, ou *kat* “vitória”, atual *katsu*). Uma vez que o *kana tsu* つ podia representar um *-t* isolado, ele foi naturalmente tomado para representar o primeiro *t* de uma sequência *tt*. A partir daí, ele se generalizou como marcador geral de geminação (FRELLESVIG, 2010).

### 3.5.5.5 Polivalência do *kana* histórico: o *hentaigana*

Durante o período histórico (apenas), em princípio qualquer caractere poderia ser cursivizado e usado como *hiragana*, ou quebrado e usado como *katakana*. Na prática, os *kana* eram normalmente tomados de um conjunto de aproximadamente 285 *kanji*, cujo uso como fonogramas já era bem estabelecido (MEXT, 2016). Esse número é significativamente maior que o número de moras possíveis em japonês, de forma que o *kana* histórico possuía alto índice de polivalência de escrita; isto é, cada mora da fala podia ser grafada de várias formas diferentes. Uma reforma ortográfica logo depois da Segunda Guerra Mundial padronizou uma grafia única para cada mora, resultando na seleção atual unívoca de 48 caracteres *hiragana* e outros 48 equivalentes em *katakana*. Os grafemas alternativos, abandonados por essa reforma, foram retroativamente denominados *hentaigana* (“*kana* irregular”). Por exemplo, os seguintes quatro *kanji* foram usados historicamente como *hiragana* para o som /i/:

Tabela 38 – Caracteres *kana* históricos para o som /i/

Forma <i>kanji</i>	Cursivização <i>kana</i>
以	𠄎
意	𠄎
伊	伊
移	𠄎

Destes, o primeiro 以 foi escolhido para a unificação moderna, na forma do *hi-*

*ragana* らがな, e os outros hoje são obsoletos.<sup>72</sup> Para listagens completas de *hentaigana*, ver (NINJAL, 2016; MEXT, 2016).

### 3.5.6 A escrita mista: *kanji-kana majiribun*, flexões e *okurigana*

Quando os fonogramas *kanji* começaram a ser usados para grafar a língua japonesa, eles eram especialmente úteis para flexões e partículas gramaticais (sem sentido referencial), cuja representação morfográfica seria confusa. Disso surgiu o princípio da **escrita mista**: grafar o conteúdo lexical (incluindo raízes e bases) morfograficamente, e as flexões e sufixos gramaticais fonograficamente. Nos editos imperiais *senmyō*, os fonogramas eram diferenciados dos demais *kanji* pelo tamanho menor. Uma vez que a notação *kana* estava estabelecida, ela substituiu os fonogramas *kanji* nesse papel:

(3.35) 行く 河 の 流れ は  
*yuk-u kawa no nagar-e wa*  
 vai-ADN rio GEN fluir-NMLZ TOP  
 “O fluxo do rio que corre...”

Observe que os morfemas de conteúdo lexical *yuk-* “ir”, *kawa* “rio” e *nagar-* “fluxo” estão escritos morfograficamente em *kanji*, que denotam seus significados; enquanto que partículas gramaticais como *no* e *wa* são grafadas fonograficamente, em *kana*.

Observe também que as flexões dos verbos são grafadas em *kana* afixados a base, que é em *kanji*. Mas há uma complicação: note a diferença entre a separação morfológica, *yuk-u*, e ortográfica, ⟨*yu.ku*⟩. O mesmo ocorre em *nagar-e*, grafado ⟨*naga.re*⟩. Isso acontece por limitações da escrita *kana*. Como os caracteres são moraicos, eles não permitem representar consoantes e vogais isoladamente. Não é possível escrever um *k* isolado para representar a raiz *yuk-*, ou um *r-* para escrever *nagar-*. Por isso, quando a base termina em consoante, a consoante acaba sendo grafada junto com o sufixo, sem fazer parte da leitura do *kanji* em si.

Note ainda que os *hiragana* são visualmente distintos dos *kanji*, por serem escritos com poucos traços e em forma cursiva (mesmo em fontes tipográficas, que mantém a fluidez curvilínea do *hiragana*). Os *katakana* também se distinguem do *kanji* pelo baixo número de traços, criando maior espaço branco. Isso serve como pista adicional para o leitor; embora o texto japonês não use espaços, a alternância de bases, afixos e partículas é refletida em uma variação visual na densidade e formato dos caracteres. Kess e Miyamoto (1999) (pp. 206–207) reveem evidência experimental que os textos misturando *kanji* e *kana* são lidos mais rapidamente, devido em parte ao fato de que a presença do

<sup>72</sup> Note porém que o *katakana* moderno, ㇰ, é baseado no *kanji* da segunda linha, 伊. O *katakana* também teve variantes pré-modernas hoje obsoletas (e.g. 子 usado para representar o som *ne*, em adição ao atual ㇰ); para uma tabela das variantes antigas de *katakana*, ver Seeley (1991, apêndice 4).

*kanji* age como indicador visual das zonas de conteúdo lexical.

Em japonês, a escrita mista de *kanji* e *kana* chama-se, literalmente, *kanji-kana majiribun*. A prática de afixar flexões em *kana* a uma base em *kanji* é chamada *okurigana* (“*kana* acompanhante”). O *okurigana* tem o efeito secundário de auxiliar na seleção de leituras *on/kun*, reduzindo a polivalência de leitura dos caracteres; como apenas os morfemas nativos são flexionados, a presença de *okurigana* seguindo um *kanji* indica que a leitura deste provavelmente será *kun*.

### 3.5.7 A escrita em paralelo: as glosas *furigana*

Nós vimos que a escrita *katakana* advém da prática de glosas: guias de leitura ou traduções escritos adjacentes aos *kanji*. A prática sobreviveu até hoje na forma do chamado *furigana* (“*kana* salpicado”) ou *rubi* (de *ruby*, um termo tipográfico inglês para fonte em tamanho pequeno). Se acrescentarmos *furigana* ao exemplo 3.35 acima, teremos:

(3.36) ゆく      かわ      なが  
 行く   河   の   流れ   は  
*yuk-u    kawa   no   nagar-e   wa*  
 vai-ADN   rio   GEN   fluir-NMLZ   TOP  
 “O fluxo do rio que corre...”

Observe os *kana* em tamanho pequeno sobre os morfogramas *kanji*. Como eles são fonogramas, a estrutura é algo como “<sup>rio</sup>河 que <sup>corr</sup>行e”, cercado morfogramas de fonogramas paralelos (acima) e sequenciais (sufixos e palavras gramaticais). O *furigana* é portanto uma escrita redundante, que representa um mesmo morfema simultaneamente de duas formas: fonografia e a morfografia.

As glosas *furigana* podem desempenhar várias funções. Em primeiro lugar, elas eliminam a necessidade de ter memorizada a leitura do *kanji*; por isso, são amplamente utilizadas em material para público em letramento, e.g. livros infantojuvenis e para aprendizes estrangeiros. Pelo mesmo motivo, elas também permitem o uso de caracteres raros ou obscuros; sempre que uma autora estima que um caractere seria difícil de ler para parte da audiência, ela pode lançar mão do *furigana*. Além disso, quando um *kanji* tem várias leituras em potencial, o uso de *furigana* especifica qual delas a autora tinha em mente, o que pode ser necessário em alguns casos; por exemplo, 一日 pode representar tanto “um dia” *ichinichi* quanto “dia primeiro do mês”, *tsuitachi*, e caso o contexto não torne óbvio qual das palavras a escritora deseja representar, as glosas permitem selecioná-la explicitamente. Um recurso estilístico muito produtivo é escrever, na glosa *furigana*, uma palavra que não é uma leitura normal dos *kanji*; discutimos isso na seção 5.2.

### 3.5.8 Kanji, polivalência e ideografia

Na seção 3.3.4, conhecemos a crítica da noção de “ideograma”: o caractere chinês não é um ideograma, pois ele não representa ideias abstratas, mas sim morfemas de uma língua. Um exemplo disso são as palavras chinesas *quǎn* e *gǒu*; ambas significam “cachorro”, mas como seus significantes são diferentes, cada uma é grafada com um *hànzì* distinto: 犬 e 狗.

Mas isso só é verdade no chinês, onde a polivalência é muito baixa (em geral os *hànzì* só têm uma leitura possível; e quando têm duas, na maioria dos casos é um alomorfe do mesmo morfema). E no caso do japonês, onde, como vimos, um mesmo caractere pode ter inúmeras leituras *on* e *kun*? Por exemplo, a palavra nativa para “cachorro” é *inu*; mas o léxico também conta com o morfema sinítico *ken*, cognato de *quǎn*, presente em palavras como *banken* “cão-de-guarda”. E tanto *inu* quanto *ken* são grafadas pelo mesmo caractere: 犬. Não seria possível dizer então que 犬 representa a “ideia” de cachorro/cão, seja ela realizada pelo morfema *inu*, seja por *ken*?

De modo geral, o japonês agrupa diversos morfemas em uma mesma grafia; em nossa contagem, a escola japonesa ensina mais 4400 leituras para 2136 caracteres, de forma que a média é aproximadamente 2 leituras.<sup>73</sup> O caso típico possui um par *on/kun*; casos extremos podem chegar a 12 leituras, como 生 “vida”. Além disso, os vários morfemas agrupados em um caractere quase sempre partilham um sentido comum. Isso é em parte um acidente histórico, porque as diversas leituras *on* são variações linguísticas da mesma palavra chinesa, e as leituras *kun* são originalmente glosas ou traduções dessa palavra. De todo modo, o resultado é que, sistematicamente, um *kanji* serve como representação visual de um campo de sentido—o que se reflete, inclusive, na organização mental do léxico da língua (JOYCE, 2004). O autor já ouviu testemunho, por exemplo, de um nativo que errou o nome de um novo conhecido, porque o que veio à mente na hora de chamá-lo foi outra leitura do mesmo *kanji*—tudo em uma conversa falada, sem troca de cartões de visita ou outro estímulo visual.

Nesse sentido limitado, o *kanji* japonês apresenta algum grau de ideografia. É preciso ter em mente, porém, que isso ainda está longe do tipo de ideograma imaginado por Fenollosa ou pelos antigos missionários. Nem todo morfema de mesmo sentido será grafado com o mesmo caractere—o japonês tem *kō* 狗 “cão”, cognato de *gǒu*, distinto de *ken/inu* 犬—e a leitura de um texto japonês é um ato linguístico, que envolve necessariamente a decodificação em significantes (uma sequência de fonemas), não uma álgebra abstrata de figuras que representam visualmente cenas da natureza.

Os mesmo fatores que geraram essa ideografia limitada também fizeram com que o japonês quebrasse completamente a relação de um-para-um entre caracteres e

<sup>73</sup> Cf. nosso programa JovoDB, nos arquivos desta dissertação, e capítulo 4 para mais dados.

sílabas, tão forte no chinês. A regularidade da morfografia também saiu enfraquecida, com bastante polivalência, plurimorfemia, e palavras compostas; contudo, as leituras mais frequentes são leituras *on* regulares, que seguem o princípio morfográfico simples (cf. capítulo 4). Mesmo com todas as irregularidades e recursos adicionais, a morfografia continua sendo a base da escrita em *kanji*, não a logografia ou ideografia (JOYCE, 2011).

#### 3.5.8.1 Mesmo som, outra nuance: a ideografia do *dōon-igi*

Na seção 3.5.4.1, vimos que palavras como “lábio” ou “trovão” são monomorfêmicas em chinês (*léi*, *chún*), razão pela qual são escritas com um só morfograma (雷 *léi*, 唇 *chún*); porém, sua tradução *kun* em japonês exige vários morfemas, o que resulta em um caractere representando uma sequência de morfemas (雷 *kami-nari* “voz dos deuses”, 唇 *kuchi-biru* “borda da boca”). Assim, a diferença na estrutura morfológica das línguas é responsável por características do uso complexo do *kanji* em japonês; no caso, por polivalência morfográfica.

Além de diferir no número de morfemas, as línguas também diferem nos vários sentidos associados a cada um, de forma que nem sempre é possível traduzi-los por equivalentes unívocos. Um fenômeno interessante ocorre quando uma palavra em japonês equivale a várias palavras chinesas distintas. Por exemplo, o verbo japonês *au* tem o sentidos de “encontrar; unir, combinar”. O primeiro sentido seria traduzido em chinês como *hé* 合; o restante, como *huì* 会—verbos distintos, e escritos com caracteres distintos. Como resultado, criou-se em japonês uma norma ortográfica: sempre que *au* tem o sentido de “encontrar”, é escrito como 会 ㄅ, e quando tem o sentido de “unir”, como 合 ㄅ. Esta grafia é dita de “mesmo som e sentido diferente” (*dōon-igi* 同音異義). Muitas vezes o termo é traduzido simplesmente por “homófono”, mas note que esses homófonos na verdade são um tipo de heterografia (sec. 3.2.2.3.2) que introduz distinções de nuance ausentes da fala. No caso do verbo *au*, a heterografia acabou especializando-se ainda mais:

Tabela 39 – *Dōon-igi*: distinções de nuance ausentes da fala

Verbo japonês	Nuance	Tradução para chinês	Ortografia <i>dōon-igi</i>
au	“combinar”	合 <i>hé</i>	合う <i>au</i>
	“encontrar”	会 <i>hùi</i>	会う <i>au</i>
	“encontrar por acaso; deparar-se com um amigo”	逢 <i>féng</i>	逢う <i>au</i>
	“encontrar infortúnio; dar com o desastre”	遭 <i>zāo</i>	遭う <i>au</i>

Há ainda o verbo *yù* 遇 “encontrar, ter a oportunidade de encontrar”, que gerou a grafia japonesa 遇う *au*, sem regra clara de uso. Tais distinções de nuance são relativamente comuns no *kanji* japonês, mas podem ser difíceis de manter na prática; “erros” de ortografia heterográfica não são raros. Neste exemplo, a distinção básica entre 合う “combinar” e 会う “encontrar” é a mais estável, seguida talvez pela nuance negativa de 遭う.

Um caso muito interessante é o do verbo *niou* “cheirar (intransitivo), exalar cheiro”. Ele é grafado heterograficamente como 匂う *niou*, quando se refere a aromas agradáveis, ou como 臭う *niou*, para mau cheiro. Seria de se esperar que os dois caracteres correspondessem originalmente a verbos chineses distintos, um significando “ser fragrante”, o outro “feder”. E de fato, 臭 corresponde ao verbo *xìu* “feder”; porém, 匂 não existe em chinês. Trata-se de um caractere criado no Japão (*kokuji*) especificamente para grafar o verbo *niou*—ou, mais precisamente, para grafar a nuance positiva de *niou*. Podemos supor que os escribas japoneses sentiram ser inadequado usar 臭, chinês “feder”, para representar *niou* quando sua acepção é positiva; porém, eles também não usaram um caractere chinês como 香 *xiāng* “perfumado; aromático” (que foi tomado para o japonês *kaoru*), preferindo criar uma nova notação que fizesse a contraparte a 臭 (HASHIMOTO, 2016). Isso demonstra um domínio bastante sofisticado da heterografia sino-japonesa, e também que os leitores sentem os efeitos da notação semântica, não apenas fonográfica. Cf. seção 5.1 para exemplos de como esse recurso é explorado.

### 3.5.9 Sistemas mistos semelhantes ao japonês

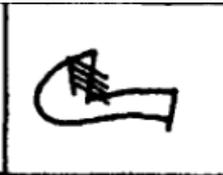
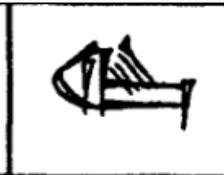
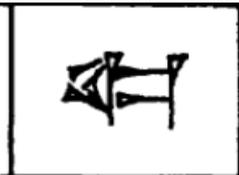
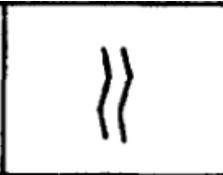
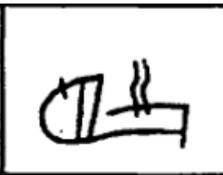
Vimos neste capítulo que a escrita japonesa moderna caracteriza-se como mista; isto é, ela é dotada de alguns caracteres para uso fonográfico (*kana*), e outros para uso morfográfico (*kanji*). Os caracteres morfográficos foram importados de uma outra língua, o chinês literário, que permaneceu sendo usada como escrita culta de prestígio.

Isso resultou em um grande influxo de empréstimos, formando uma camada lexical na língua nativa (os *kango*, “palavras chinesas”). O mesmo fato resultou também em uma dualidade na leitura morfográfica: cada caractere pode representar um empréstimo da língua original (leitura *on*), ou uma tradução na língua nativa (leitura *kun*). Por exemplo, *san* “montanha” é um empréstimo do chinês, e pode ser representado diretamente como seria em chinês, pelo caractere 山; além disso, o mesmo caractere foi tomado para representar a palavra nativa *yama*, tradução japonesa de *san*.

Tudo isso pode parecer complexo e exótico; porém, a escrita japonesa não é de forma alguma singular ou especial. Processos paralelos ocorreram na Coreia e Vietnã—países que, como o Japão, falam línguas não-siníticas, mas adquiriram a escrita a partir do chinês. Em ambos os casos também foram criados equivalentes locais à distinção *on/kun*, notações fonográficas, um estrato de empréstimos chineses, etc. Por exemplo, traduções em coreano, chamadas *hu-neum* 훈음, foram atribuídas a cada *kanji* (*hanja*) no antigo sistema chamado *idu* 吏讀/이두. Porém, a Coreia e o Vietnã terminaram por adotar a fonografia completa e abandonar os morfogramas chineses, fazendo do japonês a única escrita mista ainda em uso.

A escrita mista tampouco está restrita à área de influência do chinês literário. Considere a escrita acádia, derivada da suméria. Os símbolos cuneiformes sumérios, datados ca. 4000 AC (antecedendo em muito a escrita chinesa), são morfogramas de estruturas bastante semelhantes às que vimos na seção 3.3:

Tabela 40 – Alguns morfogramas da escrita cuneiforme (ROGERS, 2004, pp. 88–)

	Evolução do símbolo cuneiforme				Leituras em Sumério
1					/sag/ “cabeça”
2					/ka/ “boca” → /dug/ “falar”
3					/a/ “água” → /a/ “em”
4					/naḡ/ “beber”

Da esquerda para a direita, vemos como os caracteres sumérios eram originalmente pictogramas que foram simplificados para a facilidade de escrita, exatamente como os pictogramas chineses ou *xiàngxíng* (*shōkei*; cf. sec. 3.3.2.4). O caractere /sag/ “cabeça” gerou o caractere /ka/ “boca” por meio de um símbolo indicativo, como os *zhǐ shì* (*shiji*); posteriormente esse caractere também adquiriu o sentido de “falar”, /dug/. O caractere /a/, “água”, foi tomado pelo som para representar a preposição /a/ “em”, tal qual os caracteres de rebus ou *jiǎjiè* (*kashaku*). O caractere /naḡ/ é uma associação semântica, paralela aos *huiyì* (*kai'i*) chineses.

Por volta de 2350 AC, a civilização acádia conquistou a suméria. A língua suméria é isolada, sem parentes conhecidas, enquanto que o acádio é uma língua semítica, de mesma origem que o hebraico ou o árabe. Os acádios tomaram para si a forma de vida da civilização suméria, e, a despeito da diferença estrutural entre as línguas, adotaram o sumério escrito como língua culta de prestígio. Eventualmente os caracteres sumérios foram adaptados para escrever também a língua acádia, que recebeu por este meio um grande influxo de empréstimos sumérios. Os paralelos com o japonês devem estar claros. Desta forma, por exemplo, o caractere /sag/ “cabeça” tornou-se polivalente, podendo representar tanto este estrangeirismo sumério (um equivalente da leitura *on*), ou então a palavra nativa para “cabeça”, /re:sum/ (leitura *kun* acádia). O símbolo para “terra” podia representar o sumério (*on*) /ki/ ou o acádio (*kun*) /ersetu/ (ROGERS, 2004,

pp. 88–90).

Casos como o chinês/coreano e o sumério/acádio demonstram que escritas morfológicas mistas, como o japonês, aparecerão sempre que as condições históricas as favoreçam. O requisito básico parece ser a situação de diglossia, e uma escrita morfológica para a língua estrangeira de prestígio.



## 4 Medindo o grau de foneticidade do *kanji*

Neste capítulo, consideraremos a análise feita em DeFrancis (1984) (capítulo 6), que conclui que a escrita chinesa seria “66% fonética”. Analisaremos como esse número foi calculado, o que ele significa, e suas limitações. Depois, apresentaremos uma medida semelhante da foneticidade do *kanji* japonês.

### 4.0.10 A proposta de DeFrancis: Seria a morfografia uma fonografia?

Dentro dos inúmeros debates envolvendo o tema, uma importante contribuição foi a do sinologista John DeFrancis, que argumentou que “é muito justificado considerar o sistema chinês como basicamente—isto é, mais do que qualquer outra coisa—um sistema fonético de escrita” (DEFRANCIS, 1984, 111). A idéia depois seria ampliada em seu *Visible Speech: The Diverse Oneness of Writing Systems* (1989), onde o autor argumenta que toda forma de escrita é, necessariamente, fonográfica. Em sua visão, a morfografia é simplesmente uma fonografia mal-organizada, na qual a imperfeição da transcrição fonológica é compensada por um subsistema auxiliar, secundário, de pistas semânticas (op.cit, p. 253).

Esta proposta, quiçá algo ousada, pode parecer surpreendente quando consideramos as características mais salientes da morfografia, como os que vimos à seção 3.3 (página 69). Para entender o argumento do autor, vamos considerar novamente o sistema de componentes dos caracteres chineses.<sup>1</sup>

O primeiro *insight* de DeFrancis é tratar o conjunto de componentes fonéticos como um *silabário*—como um subsistema fonográfico, tal qual o *kana* japonês ou a escrita cherokee (cf. secs. 3.3.2.3 e 3.5.5). É verdade que a idéia de um “silabário” nunca fez parte da tradição chinesa dos *hànzì*; não há listagens canônicas de componentes fonéticos, e tampouco foram esses componentes categorizados em dicionários ou empregados como cabeçalhos de seção, como os *bùshǒu* semânticos.<sup>2</sup> Muitas sílabas possíveis na língua não têm grafos próprios, e, ao mesmo tempo, há muitos grafos redundantes para representar uma mesma sílaba (polivalência de escrita). DeFrancis argumenta que um sistema de escrita fonográfico imperfeito não deixa de ser fonográfico, levanta como exemplos silabários como o sumério, que é igualmente não-padronizado, e o Yi, que é igualmente numeroso (com 819 símbolos). Além disso, ortografias como as do inglês ou francês demonstram que a correspondência com o som não precisa ser perfeita para que um sistema fonográfico seja utilizável.

<sup>1</sup> O que se segue é um resumo de DeFrancis (1984), cap. 6.

<sup>2</sup> Ao menos não até a era moderna; Karlgren (1923) organizou os caracteres em termos de famílias de componentes fonéticos, e diversos filólogos posteriores seguiram seu exemplo.

Karlgren (1923), o pioneiro da reconstrução histórica do chinês antigo, argumentou que 90% dos caracteres foram formados pela estrutura fono-semântica.<sup>3</sup> Este número, porém, considera o princípio de criação dos caracteres, não os componentes que são efetivamente úteis enquanto pistas fonéticas no sistema atual (do ponto de vista sincrônico). Além disso, ele não leva em conta o fato que a maioria dos caracteres raramente é usada na prática; é preciso balancear as medidas pela frequência dos caracteres. DeFrancis concede essas limitações, e se dedica então a encontrar uma cifra mais representativa. Em primeiro lugar, é preciso decidir qual é exatamente o conjunto de símbolos do “silabário chinês”. O autor cita propostas de Karlgren, que identificou 1260 componentes fonéticos do ponto de vista da fonologia histórica; e de Zhou Youguang, que listou 1348 fonéticos no sistema moderno de caracteres simplificados da China comunista. O autor opta por trabalhar com o conjunto de Soothill (1908), missionário que identificou 895 componentes fonéticos básicos. DeFrancis chama este conjunto de “silabário Soothill”.

O próximo passo é estimar quantos caracteres efetivamente em uso têm sua leitura previsível a partir do silabário Soothill. DeFrancis nota em primeiro lugar que a maioria dos componentes fonéticos são livres, isto é, ocorrem como caracteres independentes (cf. tabela 12 à página 74). Estes são considerados foneticamente previsíveis, por definição do “silabário”. Caracteres compostos que incluem um componente do silabário são considerados “foneticamente determinados” caso possuam mais de 50% dos fonemas segmentais em comum (o raciocínio sendo que esse limite é suficiente para agir como pista mnemônica). Os demais casos são considerados não-fonéticos.

Seguindo tal critério, DeFrancis examina um conjunto de 4719 caracteres de uso frequente, tabelados por Chen Heqin a partir textos reais. O autor faz 5 amostras de 100 caracteres cada, sendo cada amostra retirada de uma faixa de frequência diferente. No total, ele estima que 14% desses 500 caracteres são grafemas livres presentes no silabário Soothill, e outros 83% são caracteres compostos; destes, 66% incluem um componente fonético útil do silabário.

É este número—66%—que DeFrancis levanta como índice de “utilidade”. O autor faz um pequeno teste com os cem primeiros caracteres de um texto literário, dos quais 58 são componentes livres do silabário, e 14 são compostos que incluem pista fonética útil, totalizando 72% com “aspecto fonético significativo”.

#### 4.0.11 Limitações da análise de DeFrancis

A discussão de DeFrancis foi influente no mundo da linguística internacional, e instrumental em divulgar a existência e importância do sub-componente fonográfico

<sup>3</sup> Caracteres 形聲 *xíngshēng*, “forma-e-som”, discutidos na seção 3.3.2.4.

da escrita chinesa.<sup>4</sup> Informativa e didática, a análise tem contudo algumas limitações.

Uma delas é que a proporção de 66% é baseada em uma amostra pequena, e não é bem balanceada por ocorrências (cf. seção 4.1.1.1). Algum balanceamento foi feito escolhendo amostras de faixas de frequência distintas, mas isso não é equivalente à análise balanceada por um corpus grande.

O índice também não leva em conta a consistência de cada pista fonética: isto é, considerando o conjunto de caracteres que incluem aquele componente, a proporção daqueles cuja leitura é prevista pelo componente, como fração do total. Uma pista pouco consistente tem muitas leituras diferentes atribuídas a ela, o que cria dúvidas na hora de utilizá-la, tornando-a menos útil. [Chen et al. \(2003\)](#), em revisão da literatura, concluem que a evidência experimental mostra que a consistência é um fator importante no processamento de componentes fonéticos.

Outro problema é que, embora DeFrancis inicialmente reconheça que um componente fonético também pode ser semântico, durante a análise ele afirma que o sentido dos componentes fonéticos é irrelevante; apenas o som importa (p. 104). Mas esta afirmação não se sustenta sem argumentos; casos como os da tabela 21 (seção 3.3.2.5 à página 85) parecem claramente incluir componentes duais, simultaneamente fonéticos e semânticos. Enquanto DeFrancis considera essas associações mera “etimologia popular, um passatempo favorito dos estudantes de chinês”, o lexicógrafo [Tōdō \(1965\)](#) propôs toda uma teoria dos caracteres chineses baseada no princípio dos componentes fono-semânticos duais. Ademais, [Zhou e Marslen-Wilson \(1999\)](#) mostram evidência experimental que os componentes fonéticos ativam mentalmente também seu lado semântico. É só por descartar esta possibilidade, e por minimizar a relevância do componente semântico, que o autor pode afirmar que a escrita chinesa é “fonética mais do que qualquer outra coisa”.

Finalmente, o índice de 66% é baseado no uso dos caracteres chineses para representar línguas chinesas; ele não reflete o uso das línguas não-chinesas que adaptaram esses caracteres, como o japonês. DeFrancis afirma que a escrita japonesa provavelmente seria ainda mais fonética que a chinesa, por dois motivos. O primeiro é a presença dos caracteres *kana*, fonográficos com alto grau de regularidade. Isso é indiscutível. Mas DeFrancis também diz que “a estrutura fonológica mais simples do japonês resultou na fusão de muitos componentes fonéticos que no chinês eram pronunciados diferentemente” (p. 113). É verdade que o número de fonemas e sílabas do japonês é bastante menor que o chinês, mas sem fazer uma análise própria, é impossível saber se isso ajudou ou *prejudicou* o sistema de pistas fonéticas; as fusões ocorreram no nível da fala, sem qualquer cuidado para respeitar as relações internas do sistema de escrita. Por fim, o

<sup>4</sup> Em uma busca em julho de 2016, o site Google Scholar listou 1444 recursos acadêmicos que citam os trabalhos mencionados.

argumento ignora o fato que a escrita japonesa, ao contrário da chinesa, rotineiramente atribui muitas leituras distintas ao mesmo caractere (polivalência de leitura).

## 4.1 O caso do japonês

No tipo de representação escrita analisada por DeFrancis, cada caractere chinês representava uma sílaba, que também era um morfema. Isto é o caso comum dos caracteres chineses, a ponto de ser assumido como regra.<sup>5</sup> Como discutido extensivamente na seção 3.3.3.3, isso deriva da própria estrutura das línguas chinesas, que têm por regra atribuir um sentido a cada sílaba (ou seja, cujos morfemas são monossilábicos).

No Japão os caracteres chineses, chamados *kanji*, são usados como parte de um sistema de escrita misto (seção 3.5). O número de caracteres distintos empregados no japonês é menor do que no chinês; o sistema de ensino japonês trabalha com um conjunto de 2136 caracteres, contra os quatro a seis mil esperados de um chinês adulto (MEXT, 2010).<sup>6</sup> Porém, como vimos, o número de leituras que precisam ser memorizadas é comparável, devido ao uso mais complexo dos *kanji* japoneses, que são polivalentes.

Dada a multiplicidade de leituras possíveis para cada caractere, e a erosão do princípio morfográfico, quanto do sistema de pistas fonéticas permanece no sistema japonês?

Como os componentes fonéticos foram desenvolvidos na China, eles servem apenas para as leituras siníticas (*on*). Além disso, a multiplicidade de leituras *on* diminui o poder de previsibilidade de cada pista; se o leitor não tem familiaridade com a palavra, ele pode inferir uma das muitas leituras quando outra é que está em jogo. Lembre-se também que em japonês as leituras rotineiramente podem ter mais de uma sílaba ou morfema, dificultando mais ainda a previsibilidade.

Um outro fator que prejudicou a utilidade dos caracteres chineses foi a simplificação (HABEIN, 2000, p. 29). Dentre os 2136 *kanji* do padrão Jōyō, 364 (ca. 17%) foram simplificados graficamente, com o objetivo de facilitar a escrita manual. O processo por vezes unificou visualmente componentes distintos, o que pode arruinar a relação entre componente fonético e pronúncia, como no exemplo:

<sup>5</sup> DeFrancis chega a afirmar que os caracteres chineses *sempre* correspondem a exatamente uma sílaba, embora nem sempre a exatamente um morfema (porque quando um morfema é polissilábico, o chinês tipicamente grafa-o com vários caracteres, um por sílaba). Na verdade há alguns caracteres de leitura polissilábica; DeFrancis provavelmente consideraria-os exceções periféricas ao sistema. Para uma visão diferente, ver discussão sobre Mair (2011), Mair (2015) na seção 3.3.3.2 (página 89).

<sup>6</sup> Kess e Miyamoto (1999) nota que um leitor de jornais em japonês pode ter que lidar com aproximadamente 6000 caracteres distintos; mas esperamos que aqueles fora do padrão Jōyō normalmente virão acompanhados de glosas *furigana*, não sendo necessário saber sua leitura de cor.

Tabela 41 – Confusão de componentes fonéticos devido à simplificação gráfica

Original		Simplificado	
Caractere	Fonético	Caractere	Fonético
私 <i>shi</i> “eu”	𠂇 <i>shi</i>	私 <i>shi</i>	𠂇 <i>shi/futsu</i>
拂 <i>futsu; hitsu; hotsu</i> “varrer, limpar”	弗 <i>futsu</i>	𠂇 <i>futsu; hitsu; hotsu</i>	𠂇 <i>shi/futsu</i>
佛 <i>futsu; butsu</i> “Buda; França”	弗 <i>futsu</i>	𠂇 <i>futsu; butsu</i>	𠂇 <i>shi/futsu</i>

Observe que, no esquema original, a pronúncia *futsu* (e relacionadas) era indicada pelo componente fonético 弗; mas, quando esse componente foi simplificado para 𠂇, tornou-se idêntico ao de 私, que sugere *shi*. No sistema resultante, o valor de 𠂇 como pista fonética acabou prejudicado pela polivalência de leituras.

A simplificação tem um potencial de confusão ainda maior por não ser consistente (diferentemente da simplificação implementada na China comunista, que foi consistente e completa). Por exemplo, embora 黃 *kō* tenha sido simplificado para 𠂇 em 廣 > 𠂇, ele permaneceu inalterado em 橫 *ō*; e embora 弗 *futsu* tenha sido simplificado para 𠂇 em 拂 > 𠂇, ele permaneceu inalterado em 沸 *futsu*.

Por outro lado, os japoneses usam um conjunto menor de caracteres; as leituras tipo *on* são as mais frequentes (cf. seção 4.2.1), e estas raramente ultrapassam duas sílabas; e, como DeFrancis argumentou, o número de fonemas é bem menor. Levando tudo isso em conta, o *kanji* japonês é mais ou menos fonético que o *hànzì* chinês? A questão não tem resposta fácil, e é digna de investigação.

Para a análise que apresentamos aqui, buscamos empregar técnicas estatísticas e computacionais, a fim de superar algumas das limitações da análise manual (cf. seção 4.0.11). Adotamos um ponto de vista puramente sincrônico, sem levar em conta dados etimológicos ou da história dos caracteres, pois queremos investigar quanta informação ainda resta no sistema para o leitor moderno. E, como queremos saber os componentes fonéticos de fato em uso, levamos em conta a frequência dos *kanji* em um corpus real.

#### 4.1.0.1 Funções dos componentes fonéticos

Podemos distinguir pelo menos três utilidades distintas para os componentes fonéticos, em diferentes tarefas:

1. **Escrita.** A falante quer escrever um morfema, e precisa recuperar o caractere da memória (*recall*). Ou seja, ela tem acesso à leitura e ao significado, e precisa recuperar a forma visual (e gestural). Caracteres frequentes são recuperados facilmente, mas um *kanji* raro pode falhar em vir à mente. Em alguns casos, a falante começa um processo interativo, esboçando alguns componentes fonéticos e semânticos relacionados ao morfema, e conferindo se “parecem corretos”—ou seja, usando-os como ganchos para recuperar o caractere.
2. **Reconhecimento (leitura de palavra conhecida).** A falante depara-se com uma palavra escrita. Ela esqueceu como se leem aqueles *kanji*; porém, ela conhece a palavra que eles representam. Os componentes fonéticos podem servir como sugestão para, somados ao contexto, sugerir qual é a palavra.
3. **Aquisição (leitura palavra nova).** A falante depara-se com uma palavra escrita com um *kanji* que não conhece. Com base em um componente fonético, ela estima a leitura, e adquire uma nova palavra em seu léxico.

No caso da aquisição, componentes fonéticos aproximados ou inconsistentes nem sempre ajudam, e podem inclusive atrapalhar, caso a leitura prevista seja diferente da que realmente ocorre. Já no caso do reconhecimento, uma pista fonética aproximada é suficiente; o leitor pode validá-la em seu léxico mental, com a ajuda do contexto, e recuperar a leitura correta.

Nosso objetivo será medir o caso 3, reconhecimento, como a tarefa mais comum envolvendo texto escrito.

#### 4.1.0.2 Hipótese

Nossa hipótese de trabalho é que o *kanji* japonês é menos fonográfico do que DeFrancis implica. Testaremos se o grau de fonografia é baixo o suficiente para que faça sentido considerar esse subsistema como basicamente (isto é, mais do que qualquer outra coisa) morfográfico.

Adotaremos para a análise o **princípio da cautela**: em todas as variáveis não testadas, assumiremos o caso ótimo para a foneticidade. Isto é, buscaremos uma estimativa cautelosa, conservadora, para a hipótese de trabalho.

### 4.1.1 Detalhes metodológicos

#### 4.1.1.1 Tipos e ocorrências

Fazemos uso amplo da distinção entre **tipos** e **ocorrências** (*types and tokens*), conceito importante na linguística, estatística e filosofia.<sup>7</sup> Uma contagem de tipos consi-

<sup>7</sup> Cf. [Wetzel \(2014\)](#).

dera cada ocorrência apenas uma vez, sem repetições. Uma contagem de ocorrências inclui as repetições. Por exemplo, considere a expressão: “guerreiros com guerreiros”. Quantas palavras ela tem? A pergunta tem dois sentidos possíveis. Em um sentido, a resposta é duas palavras (“guerreiros” e “com”); em outro, três palavras (“guerreiros”, “com”, “guerreiros”). A primeira resposta é uma contagem de tipos; a segunda, de ocorrências.

Note que a contagem de ocorrências equivale a balancear os resultados pela frequência; ou seja, itens mais frequentes têm peso maior nesta contagem. É razoável assumir que a contagem de ocorrências indica melhor quais são os componentes fonéticos efetivamente úteis.

#### 4.1.1.2 Os *kanji* de uso geral: *Jōyō Kanji*

Como o conjunto de caracteres *kanji* é aberto, é preciso decidir qual subconjunto será analisado. Optamos por limitar nossa análise aos 2136 *kanji* ensinados do currículo escolar japonês, os “*kanji* de uso geral” ou *Jōyō Kanji*.<sup>8</sup>

Há vários motivos para esta escolha. Em primeiro lugar, esta é a tarefa que é proposta para os aprendizes de escrita japonesa, sejam crianças nativas ou estrangeiros. Ou seja, estamos medindo a utilidade das pistas fonéticas para a tarefa básica de ler textos no padrão ortográfico japonês moderno.

Além disso, é prática no Japão que leituras não-*Jōyō* sejam acompanhadas de glosas fonográficas (*furigana*). Um *kanji* com glosas é uma escrita foneticamente determinada, o que afetaria nossa análise; porém, é difícil identificar as glosas nos formatos textuais da Internet, inclusive em nosso corpus. Limitar-se às leituras *Jōyō* evita este problema, já que por definição estas não precisam de glosas para o leitor adulto.

O *Jōyō* é publicado pelo Ministério da Educação (MEXT) em formato PDF, que de é difícil tratamento programático. Nós produzimos software capaz de extrair os dados do PDF e salvá-los em formato de planilha. O programa resultante, *JoyoDB*, está disponível no sítio dos arquivos desta dissertação, e foi publicado como *software* livre para uso do público, assim como os dados extraídos.

#### 4.1.1.3 O corpus e seu tratamento

Usamos como corpus o texto da Wikipedia japonesa (edição de 22 de abril de 2015).<sup>9</sup> Processamos esse corpus da seguinte forma:

1. O código de formatação foi removido através da ferramenta *wp2txt*, seguindo as instruções de Yamaguchi (2011).

<sup>8</sup> MEXT (2010).

<sup>9</sup> Wikimedia Foundation (2015)

2. O texto resultante foi analisado pelo analisador morfológico Mecab, usando o dicionário Unidic.<sup>10</sup> Este passo é necessário porque precisamos da transcrição fonológica do texto japonês, e devido à natureza da escrita japonesa (cf. seção 4.1), é impossível chegar a essa transcrição sem antes segmentar os morfemas. Extraímos do Mecab os morfemas em forma **lematizada** (ou seja, flexionados na “forma de dicionário”), bem como suas transcrições. Embora se chame “analisador morfológico”, o Mecab nos retorna palavras inteiras, quando é relevante; chamaremos as unidades genericamente de **lemas**, sejam elas morfemas ou palavras. Filtramos a análise do Mecab por lemas cuja grafia original inclua caracteres *kanji*. Os *scripts* Python que desempenham essas tarefas estão disponíveis nos arquivos desta dissertação.

3. Muitos lemas empregam mais de um *kanji* em sua grafia. Lembre-se que uma leitura de um *kanji* pode ter um número arbitrário de fonemas ou sílabas. Consequentemente, não é simples estimar qual segmento da transcrição do lema corresponde a qual *kanji*. Isso só é possível a partir de um dicionário de leituras potenciais de cada *kanji*, e caso a leitura do lema em questão seja derivável a partir destas. Produzimos uma biblioteca Python para este cálculo, chamada YOMISPLIT, que foi disponibilizada na Internet.<sup>11</sup> O YOMISPLIT faz uso do dicionário KANJIDIC, que lista as leituras dos caracteres.<sup>12</sup>

Com ajuda do YOMISPLIT, um *script* Python separa os lemas em cada *kanji* e sua leitura, tal como ocorreram no corpus. O YOMISPLIT também classifica cada leitura: se é *on*, *kun*, ou outras; e se é ou não parte do Jōyō.

4. Para esta última classificação, usamos dados diretamente da tabela em PDF do Jōyō Kanji, tal como publicada pelo governo japonês, a fim de evitar corrupções comuns em fontes secundárias. Contudo, o formato PDF é de processamento algo trabalhoso, especialmente somado ao *layout* idiossincrático da tabela. Produzimos um *software* especializado para este processamento, o JOYODB, que também foi disponibilizado ao público.<sup>13</sup>

5. Um último *script* usa a base de dados KANJIVG para descobrir quais os componentes de cada *kanji*.<sup>14</sup> Os componentes são tabelados com os caracteres e leituras com que ocorreram. As inconsistências encontradas nessa base foram reportadas a seus autores.

6. Do corpus original assim processado, extraímos uma amostra de 300.000 ocorrências de caracteres *kanji*. As medidas quantitativas, estatísticas e gráficos apresentadas a seguir foram geradas usando o ambiente R.<sup>15</sup> Todas as contagens neste trabalho

<sup>10</sup> KUDŌ (2013). Agradeço a Jim Breen, do projeto EDICT, pela orientação sobre dicionários do Mecab.

<sup>11</sup> Boiko (2015).

<sup>12</sup> Breen et al. (2016).

<sup>13</sup> Boiko (2016).

<sup>14</sup> Apel et al. (2013).

<sup>15</sup> R Core Team (2015).

foram produzidas a partir desta amostra.

#### 4.1.1.4 Fontes de erros

A segmentação e transcrição fonética do analisador morfológico Mecab não são perfeitas. Em testes, [Mori \(2011\)](#) obteve precisão de 93,69% (mas com um dicionário mais antigo).

Nossa ferramenta YOMISPLIT também não é capaz de descobrir todas as leituras regulares, de forma que a classificação de leituras como “especiais” está superestimada. Combinando os casos em que o Mecab não descobriu a leitura com os casos que o YOMISPLIT não pôde processá-las, tivemos um total de 3312 leituras não analisáveis (1,1% do total). Este número inclui:

- Leituras *jukujikun* (cf. seção 3.5.4.1). Algumas das leituras *jukujikun* são Jōyō, mas o YOMISPLIT não tem suporte a este tipo de grafia, o que significa que a estimativa da proporção de leituras Jōyō está levemente reduzida. A dificuldade maior é a polivalência de leitura; leituras *jukuji* como 一日 = *tsuitachi* competem com leituras regulares para a mesma grafia, como 一日 = *ichinichi*, e não é trivial decidir algorítmicamente qual está em uso a cada ocorrência. Porém, este tipo de leitura é de toda forma irrelevante para a análise de foneticidade, que só depende de leituras *on*.
- Leituras de nomes próprios, toponímia e outros casos particulares, também irrelevantes para nossos objetivos.
- Leituras *on* e *kun* válidas, que os algoritmos não reconheceram. As primeiras são erros que afetam nossa análise. Pressupomos que a taxa (menor que 1,1%) é pequena o suficiente para não afetar as conclusões.

Uma fonte de erros mais preocupante é a base de dados KANJI VG, que usamos para dividir os caracteres em componentes. A forma com que a base divide e classifica os componentes nem sempre é adequada aos nossos propósitos (incluindo, por exemplo, critérios etimológicos ou aproximações gráficas), e a base toda precisa de uma revisão mais cautelosa, pelo menos dos 2136 caracteres Jōyō.

#### 4.1.1.5 Outras abordagens

DeFrancis usou uma proporção simples como estimativa de foneticidade, e este trabalho segue a mesma linha. Formas mais diretas da utilidade dos componentes fonéticos incluem os testes neuro- e psicolinguísticos. Uma discussão sobre essas abordagens foge ao nosso escopo, mas [Erbaugh \(2002\)](#) traz uma coletânea interdisciplinar

sobre o estudo científico dos caracteres chineses. [Chen et al. \(2003\)](#) faz uma revisão da literatura. [Dehaene \(2009\)](#) é uma introdução em linguagem acessível à neurologia da leitura, inclusive a chinesa; e, com [Nakamura et al. \(2005\)](#), o pesquisador apresenta indícios de processamento diferenciado para *kanji* e *kana*. Evidência mais detalhada nesse sentido é apresentada por [Kess e Miyamoto \(1999\)](#), e [Handel \(2015\)](#) cita resultados recentes.

## 4.2 Descrição da amostra

### 4.2.1 Caracteres *kanji* e leituras

Uma leitura de um *kanji* é considerada de uso geral, “Jōyō”, se aquela leitura ocorrer na lista Jōyō Kanji, o padrão do sistema educativo. Cada leitura pode ser sinítica (*on*), nativa japonesa (*kun*), ou especial—categoria que inclui também leituras não analisáveis por nossas ferramentas (ver seção 4.1.1). A tabela 42 abaixo mostra as contagens e proporções de cada categoria em nossa amostra.

**Tabela 42 – Contagens de caracteres *kanji* e leituras na amostra**

	Ocorrências	Tipos
Total de <i>kanji</i> :	300 000	2 693
Total de caracteres Jōyō:	295 958	2 177
Proporção Jōyō:	0,9865	0,7861
Total de leituras:	300 000	1 516
Total de leituras Jōyō:	289 625	1 097
Proporção de leituras Jōyō:	0,9654	0,7236
Leituras Jōyō siníticas ( <i>on</i> ):	241 646	295
Leituras Jōyō japonesas ( <i>kun</i> ):	46 614	761
Outras leituras Jōyō:	1 365	41
Proporção <i>on</i> (Jōyō):	0,8343	0,2689
Proporção <i>kun</i> (Jōyō):	0,1609	0,6937

Em 907 das 300.000 ocorrências, não foi possível deduzir computacionalmente as leituras; essas leituras foram classificadas como de um único tipo, não-analisável.

Na contagem de ocorrências, a grande maioria das leituras (96,54%) foi detectavelmente Jōyō. Essa proporção alta é esperada, dado que essas são as leituras ensinadas na escola. Note que o número de ocorrências de caracteres Jōyō é maior que o de leituras Jōyō. Isso é porque alguns caracteres Jōyō possuem leituras não-Jōyō. A proporção relativamente baixa de caracteres Jōyō provavelmente se deve à natureza do texto enciclopédico, que tende a citar todo termo pelo menos uma vez, para defini-lo.

Na contagem de tipos, as leituras japonesas nativas predominam (69,37%); mas, na contagem de ocorrências, as mais frequentes são as leituras siníticas (83,43%). Isso significa que as leituras nativas são mais diversificadas, mas as leituras siníticas são mais usadas. Como dissemos antes, apenas as leituras siníticas podem ser previstas por componentes fonéticos. Isso implica que 16,57% das leituras que ocorreram foram não-fonéticas por definição—bem como 69,37% dos tipos de leitura.

#### 4.2.2 Componentes

A tabela abaixo relaciona o número de componentes aos caracteres. Contamos componentes por tipos, recursivamente, considerando o próprio caractere como um componente. Por exemplo, 口 tem um componente (ele próprio), enquanto 操 tem 5 componentes: 操,手,品,口,木. Definir o próprio caractere como um componente ajuda a detectar casos em que a leitura do caractere é empregada como componente fonético; por exemplo, 青 *sei* “azul” como pista fonética em 精 *sei* “vitalidade”.

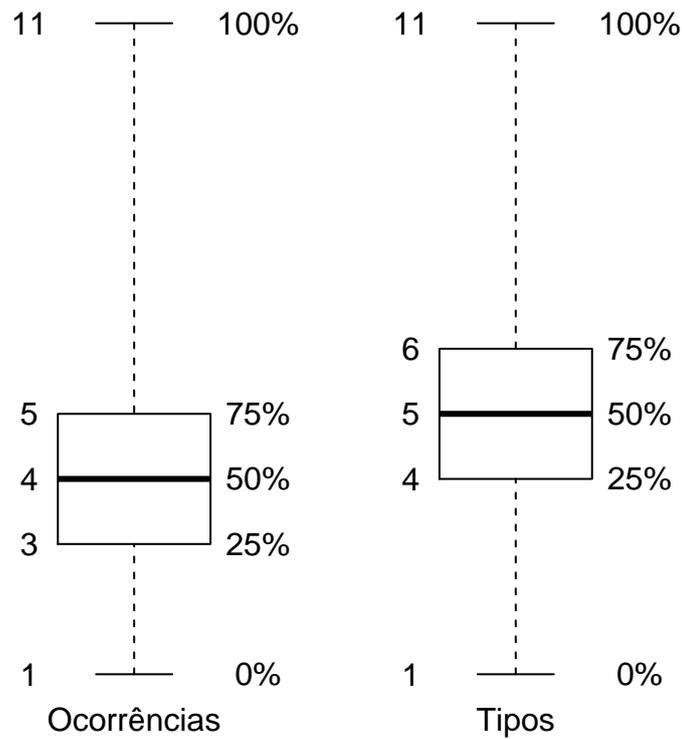
**Tabela 43 – Relação entre caracteres *kanji* e número de componentes**

	Ocorrências	Tipos
Média de componentes / <i>kanji</i> :	3,8936	5,0090
Desvio-padrão:	1,8517	1,9220

Quartis	Ocorrências					Tipos				
	0%	25%	50%	75%	100%	0%	25%	50%	75%	100%
	1	3	4	5	11	1	4	5	6	11

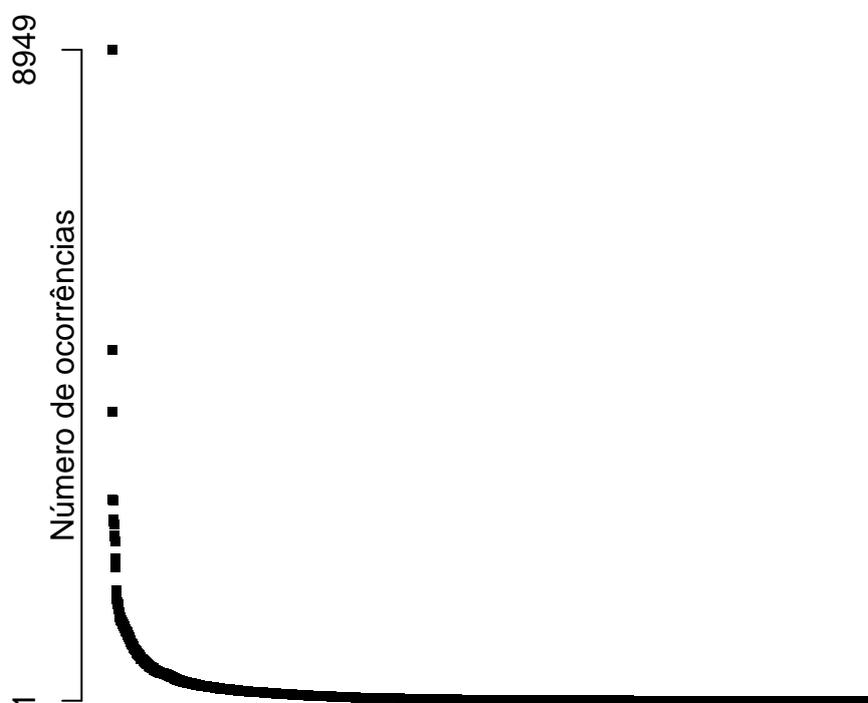
O número médio de componentes é maior na contagem de tipos. Isso quer dizer que os caracteres com pouco componentes são mais frequentes, pesando a contagem de ocorrências para baixo. A contagem dos quartis ilustra a distribuição: 25% das ocorrências têm 3 componentes ou menos, 50% 4 componentes ou menos, etc. Uma visualização da distribuição é provida abaixo no formato gráfico de caixas (*boxplot*), que deixa claro que 50% dos valores se concentram em torno das medianas  $\pm 1$ .

**Figura 6 – Distribuição do número de componentes por *kanji* na amostra, em quartis. Componentes são contados recursivamente, incluindo o próprio caractere.**



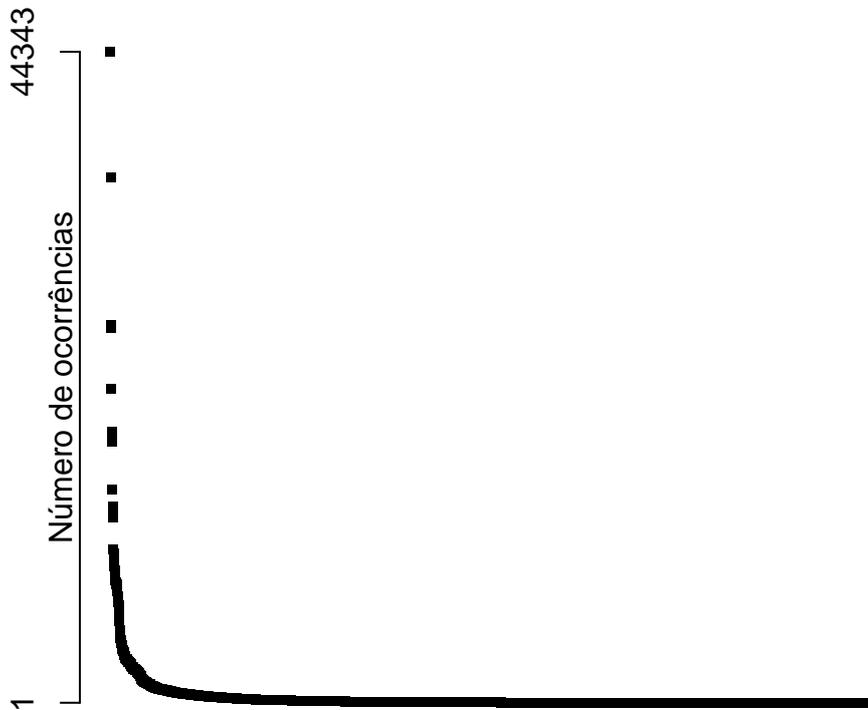
### 4.2.3 Frequências

Observe os gráficos de frequência:

Figura 7 – Frequência de cada *kanji* na amostra

Os três caracteres desproporcionalmente frequentes (visíveis como três pontos elevados no gráfico) são 年, 月, e 日—“ano”, “mês” e “dia”. Possivelmente o corpus Wikipedia menciona datas com frequência atípica, por exemplo em listagens de eventos por ano (o corpus foi limpo de metadados, tais como as datas de modificação das páginas, mas mesmo assim esses caracteres predominam). Esta frequência, porém, não altera a análise; mesmo excluindo estes três caracteres, o formato do gráfico se mantém—alguns poucos *kanji* são muito mais frequentes que a grande maioria. O mesmo ocorre com as ocorrências de cada componente:

Figura 8 – Frequência de cada componente na amostra



Deve estar claro que as duas medidas de frequência (de caracteres e de componentes) possuem um formato semelhante: o de uma distribuição logarítmica. Como em muitos fenômenos linguísticos relacionados a contagens de frequências, estas medidas parecem seguir aproximadamente a lei de Zipf, segundo a qual a frequência de um item é inversamente proporcional à sua posição na ordem de frequências (de forma que o item mais frequente ocorre aproximadamente o dobro de vezes que o segundo mais frequente, o triplo que o terceiro, etc.).<sup>16</sup> Os *kanji* não são morfemas ou parte da língua, mas distribuem-se com o mesmo tipo de organização.

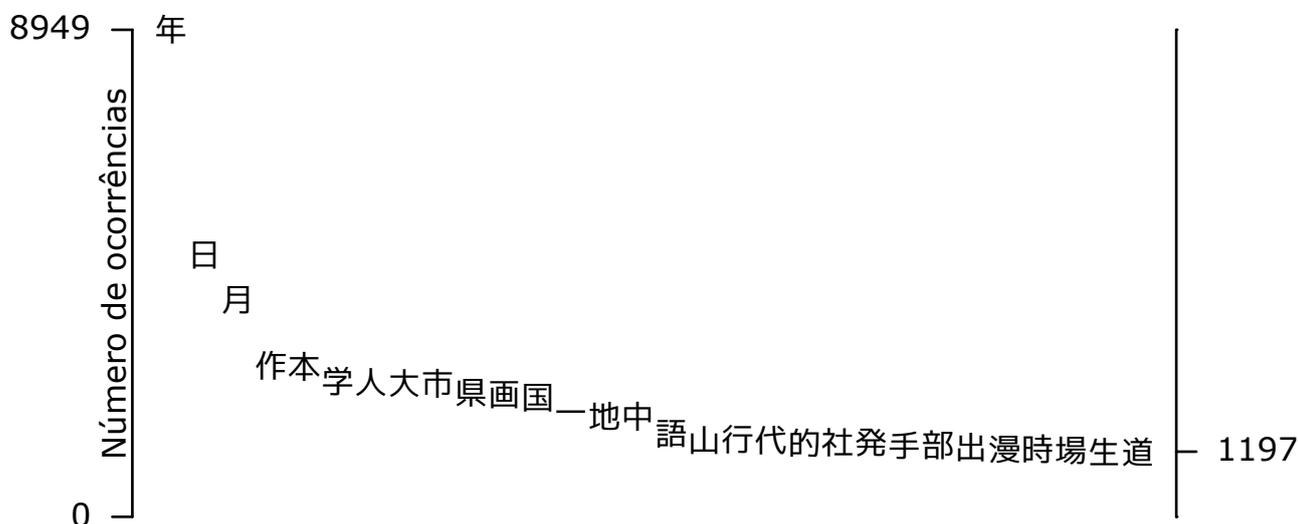
Uma consequência importante é que alguns poucos itens mais frequentes dominam a grande maior parte das ocorrências. Observe:

<sup>16</sup> Cf. e.g. [Wyllys \(1981\)](#).

Tabela 44 – Cobertura dos itens mais frequentes sobre o total de ocorrências

1% dos caracteres <i>kanji</i>	21,78% das ocorrências
10% dos caracteres <i>kanji</i>	68,14% das ocorrências
25% dos caracteres <i>kanji</i>	89,6% das ocorrências
1% dos componentes	39,84% das ocorrências
10% dos componentes	81,87% das ocorrências
25% dos componentes	94,52% das ocorrências

Figura 9 – Os 1% caracteres *kanji* mais frequentes. Juntos, estes 30 caracteres somam 21,78% do total de ocorrências na amostra. Talvez refletindo a natureza enciclopédica da amostra, caracteres que representam tempo e espaço são predominantes.



Conversamente, uma grande parcela dos tipos aparece poucas vezes: 75% dos *kanji* (na contagem de tipos) cobrem juntos apenas 10,4% das ocorrências.

## 4.3 Análise do corpus

Na seção 4.3.1 a seguir, definiremos alguns conceitos usados nesta análise, e na seção 4.3.2, listamos nossos pressupostos. Em seguida, na na seção 4.3.3 apresentamos as medidas quantitativas.

### 4.3.1 Definições

Definimos a **série** de um componente como o conjunto de caracteres *kanji* que incluem aquele componente.<sup>17</sup> Por exemplo:

<sup>17</sup> Alguns autores falam também em **família** de caracteres, com o mesmo sentido.

Tabela 45 – Exemplos de séries de componentes

Componente	Série	Tamanho da série
走	走 越 起 趣 超 徒 赴	7
青	青 情 晴 清 精 請 靜	7
乍	作 昨 搾 詐 酢	5
韋	偉 違 緯 衛 韓	5
瓜	孤 弧	2

Seguindo os critérios explicados na seção 4.1.1, estamos nos limitando ao universo dos caracteres de uso geral, Jōyō Kanji. O tamanho das séries seria maior se incluíssemos caracteres não ensinados no sistema educacional.

Definimos **vizinhança** de um par {kanji, componente} como o conjunto dos outros caracteres que incluam aquele componente; ou seja, seus “vizinhos” na série do componente. Cada kanji terá, por tanto, uma vizinhança diferente para cada componente. Considere os seguintes exemplos, comparando-os com a série 乍 na tabela 45 acima:

Tabela 46 – Exemplos de vizinhanças

Kanji	Comp.	Vizinhança	Tamanho da vizinhança
作	乍	昨 搾 詐 酢	4
作	亻	位 依 偉 促 側 俗 休...	186
昨	乍	作 搾 詐 酢	4
昨	日	日 昧 昭 春 旬 昇 晶...	229

Ressaltamos que a vizinhança de um kanji exclui o próprio kanji.

Definimos **consistência fonética** de um par {componente, leitura} como a proporção da série do componente que tem aquela leitura.<sup>18</sup> O valor numérico da consistência (ou índice de consistência) será diferente para tipos e ocorrências. Considere, por exemplo, as seguintes séries, seus caracteres e leituras (na contagem de tipos):

<sup>18</sup> Os termos “consistência” e “vizinhança” seguem Tzeng et al. (1995) e Chen et al. (2003).

Tabela 47 – Exemplos de séries e leituras

Componente	Série e leituras						
走	走	越	起	趣	超	徒	赴
	etsu otsu	ki	shu	sou	chou	to	fu
青	青	情	晴	清	精	請	静
	sei	sei	sei	sei	sei	sei	sei
	jou		shou	shou	shou	shou	jou
			shin		shin		
乍	作	昨	搾	詐	酢		
	saku sa	saku	saku	saku		sa	

Agora observe os índices de consistência para algumas leituras das séries acima (novamente na contagem de tipos):

Tabela 48 – Exemplos de consistência fonética (contagem de tipos)

Componente	Leitura	Número / Total	Índice de consistência
走	etsu	1/8	0.125
青	sei	7/15	0.47
青	shou	4/15	0.27
乍	saku	4/6	0.67
乍	sa	2/6	0.33

Definimos **melhor vizinhança fonética** de um par {*kanji*, leitura} como a sua vizinhança com maior valor de consistência. O **melhor componente fonético** de um *kanji*, ou simplesmente seu “componente fonético”, é o componente de sua melhor vizinhança fonética. Exemplo (cf. tabelas 46 e 47):

Tabela 49 – Exemplo de escolha da melhor vizinhança fonética

Kanji	Leitura	Componente	Consistência fonética	Melhor?
作	saku	イ	0	
作	saku	乍	0.67	sim
静	sei	青	0.46	sim
静	sei	争	0.33	

Na primeira linha, 作 é o único caractere de uso geral com o componente 亻 e leitura *saku*. Ou seja, nenhum *kanji* na vizinhança de {作, 亻} se lê *saku*, resultando em índice de consistência fonética 0: o componente 亻 não é uma pista fonética.

Na terceira linha, excluindo-se o *kanji* 静, restam 13 leituras na série de 青, das quais 6 são *sei*.  $6/13 = 0.46$ . Esse índice de consistência é maior que o da vizinhança via componente 争—vizinhança que possui três leituras, das quais apenas uma é *sei* (consistência = 0.33).

Enfatizamos que os índices de contagem de tipos, usados como exemplos acima, podem ser drasticamente diferentes da contagem de ocorrências, devido à distribuição logarítmica das frequências (p. 142). Ou seja, alguns componentes e leituras são extraordinariamente mais frequentes, e isso aumenta muito seus índices de consistência balanceados por ocorrência.

Note também que, mesmo que o componente  $C$  seja o melhor componente fonético de  $K_1$ , e que  $K_2$  seja seu vizinho, o componente  $C$  pode não ser um boa pista fonética no caso de  $K_2$ . Isso porque a vizinhança de  $K_1$  exclui o próprio  $K_1$ , e a vizinhança de  $K_2$  exclui  $K_2$ . Portanto, suas vizinhanças via  $C$  são diferentes, e consequentemente seus índices de consistência também diferem. Este modelo reflete a tarefa de leitura de reconhecimento (p. 135). No momento de ler um *kanji*  $K_1$ , se queremos estimar o valor de sua pista fonética, estamos pressupondo que o leitor *não* recuperou da memória a leitura do próprio  $K_1$ . Afinal, se o leitor soubesse a leitura de  $K_1$ , ele não precisaria de uma pista de leitura em primeiro lugar! O caso de máxima utilidade para uma pista fonética é quando o leitor sabe a leitura de todos os caracteres que incluem aqueles componentes, menos o  $K_1$ ; este é o caso modelado pelo conceito de “melhor vizinhança fonética”. Pelo princípio da cautela (p. 136), este é o caso que medimos.

### 4.3.2 Pressupostos da análise

A consistência fonética é uma medida estrutural e de proporção, sendo portanto diretamente relacionada à forma da análise de DeFrancis (seção 4.0.10). Experimentos psicolinguísticos mostram que a medida de consistência está implicada na tarefa de leitura, especialmente se a proporção for balanceada pela frequência (i.e., contagem de ocorrências); cf. Chen et al. (2003). Por esses motivos, escolhemos esta medida como indicador da foneticidade da escrita japonesa.

Um *kanji* pode ter vários componentes. Potencialmente nenhum deles tem valor fonético; e, mesmo que algum seja fonético, não há indicações de qual dos componentes seria a pista de leitura (há tendências gerais de acordo com a posição, mas exceções são comuns). Em termos das definições acima, um *kanji* pode ter várias vizinhanças, e não é óbvio qual delas o leitor deve tomar como pista fonética. Além disso, cada

componente pode prever leituras diferentes, e não é óbvio qual das leituras possíveis o leitor deve escolher. Novamente pelo princípio da cautela, consideramos o melhor caso do ponto de vista da fonografia. Estamos supondo um leitor que consiga escolher o melhor componente fonético para a leitura correta, graças ao seu conhecimento de léxico e compreensão do contexto, que servem para descartar leituras e pistas que não se encaixem.

Vimos na seção 4.2 (p. 140) que 16,57% dos caracteres de nossa amostra ocorreram com leituras não-siníticas (leituras não-*on*), que por definição não empregam componentes fonéticos. Porém, ao se deparar com uma ocorrência de um *kanji*, nem sempre é simples decidir se a leitura utilizada é sinítica ou não. Novamente, assumimos o melhor caso, ou seja, que o leitor consegue usar seu conhecimento lexical e contextual para decidir qual o tipo de leitura.

Com essas suposições, estamos medindo uma estimativa conservadora para a foneticidade da escrita japonesa; trata-se de uma estimativa máxima, um limite superior. Em particular, note que as pistas com baixo índice de consistência devem ser menos úteis na tarefa de aquisição de *kanji* (p. 135), já que irregularidades levam a previsões incorretas de leitura (o que é pior do que a ausência de pistas fonéticas).

### 4.3.3 Medindo a consistência das pistas fonéticas dos *kanji*

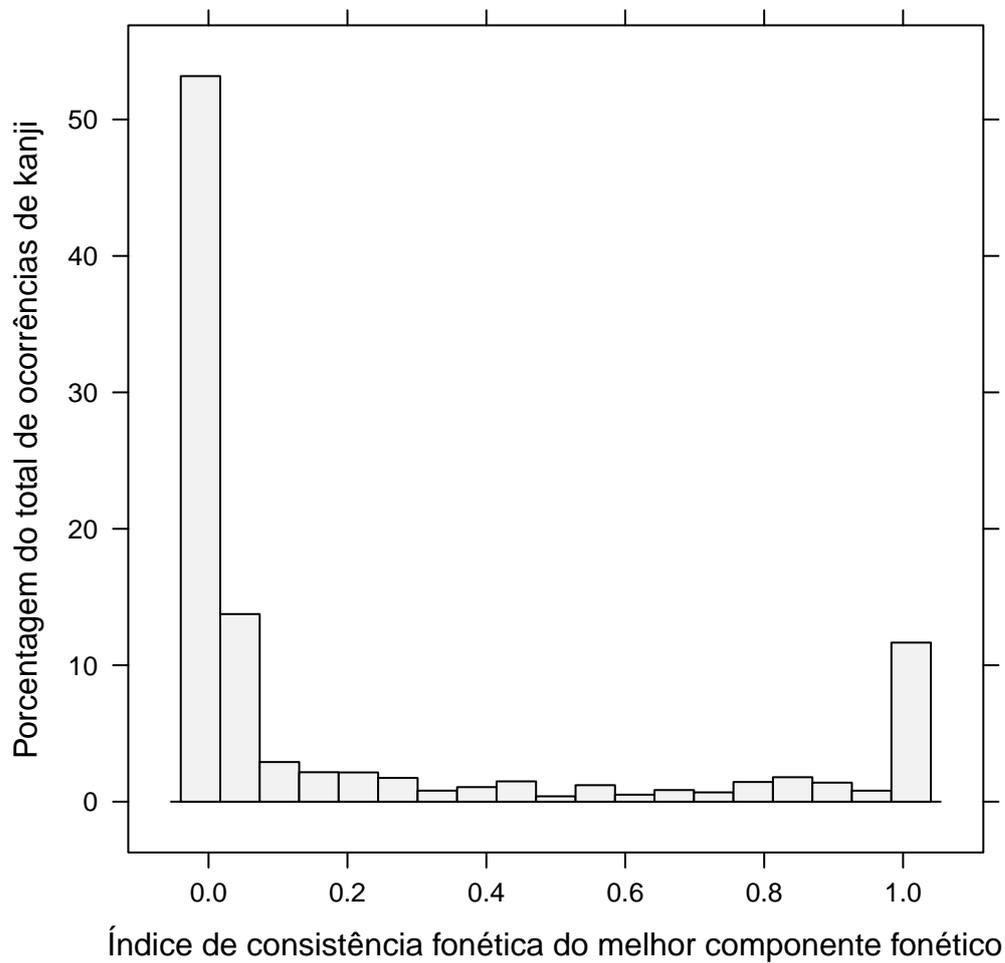
Para cada ocorrência de leitura sinítica (*on*) na amostra, calculamos o melhor componente fonético e seu índice de consistência. A tabela abaixo resume os resultados.

**Tabela 50 – Índices de consistência fonética das leituras siníticas (*on*) que ocorreram na amostra**

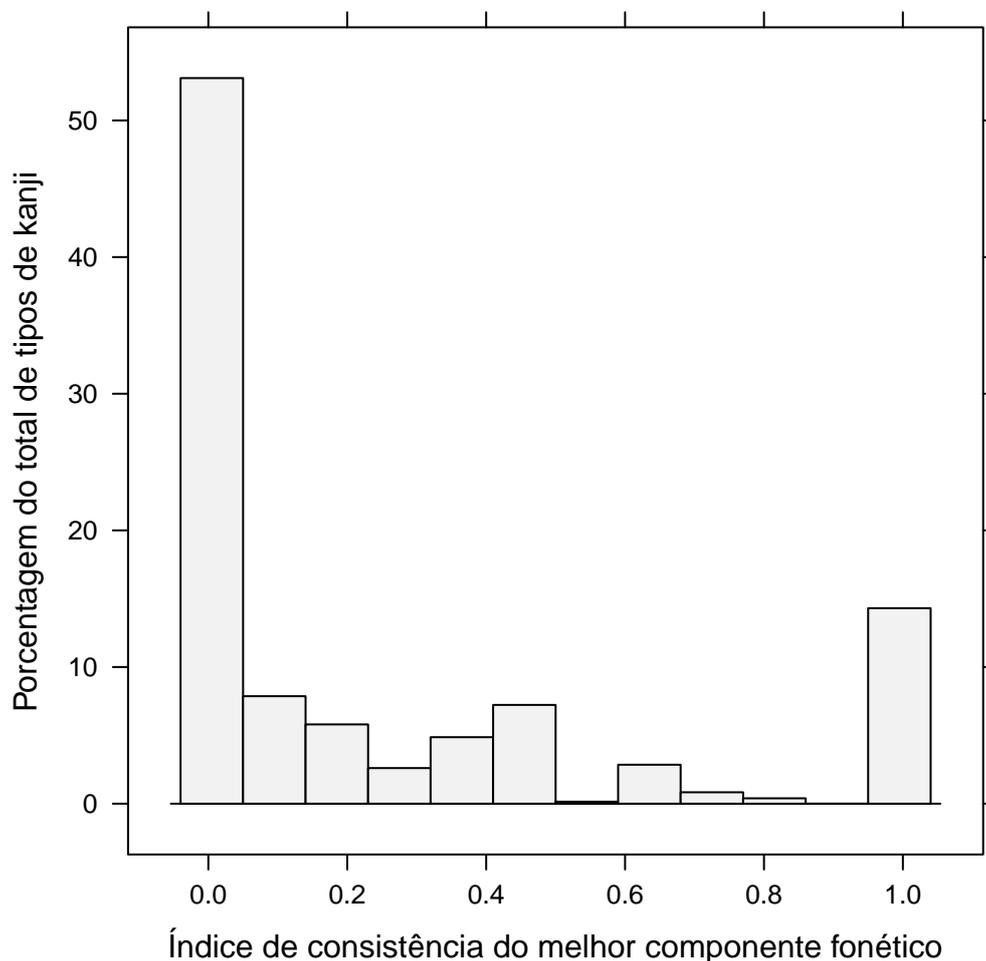
	Ocorrências					Tipos				
Consistência média:	0,2218					0,2542				
Desvio-padrão	0,3623					0,3553				
Quartis	Ocorrências					Tipos				
	0%	25%	50%	75%	100%	0%	25%	50%	75%	100%
	0	0	0,0111	0,2812	1	0	0,0033	0,0374	0,5	1

A contagem dos quartis ilustra a distribuição: 50% das ocorrências de *kanji* têm índice de consistência fonética de 0,0111 ou menos; 75% têm 0,2812 ou menos, etc. 11,24% das ocorrências de *kanji* tiveram consistência fonética 1, ou seja, eram completamente previsíveis a partir do componente fonético. Os índices são mais altos na contagem de tipos, o que significa que as leituras menos previsíveis tendem a ocorrer com mais frequência. As distribuições podem ser visualizadas nos histogramas:

Figura 10 – Histograma da distribuição do melhor índice de consistência fonética para ocorrências de *kanji* na amostra



**Figura 11 – Histograma da distribuição do melhor índice de consistência fonética para tipos de *kanji* na amostra**



Observe o formato em U dos gráficos, com a ponta direita mais baixa. Uma pequena proporção da amostra, pouco mais de 10%, tem sua leitura completamente previsível (extremidade direita); mais da metade é absolutamente imprevisível foneticamente (extremidade esquerda); e o restante distribui-se quase que uniformemente pelo espaço de possibilidades.

#### 4.4 Discussão e perspectivas

Buscamos analisar uma amostra grande de um texto japonês contemporâneo, e avaliar quanta informação fonética há nele para um leitor educado no sistema escolar padrão. Limitamo-nos a uma estimativa cautelosa, no caso ótimo para a previsibilidade: consideramos apenas a tarefa de recuperação de palavras conhecidas, assumindo que o leitor sabe estimar quais leituras no texto são siníticas (*on*), quais componentes são fonéticos, e qual das possíveis leituras é a que se encaixa. Assumimos ainda que o lei-

tor tem familiaridade com os outros *kanji* que partilham o mesmo componente de cada ocorrência. Mesmo nesta situação idealizada, o grau de foneticidade medido foi comparativamente baixo. 16,57% das ocorrências não eram *on*, e das restantes, a maior parte possuía índices de consistência fonética menores que 0,3. Não há como estabelecer um limite preciso para o que conta como “componente fonético”, mas se escolhermos cautelosamente um critério de consistência igual ou maior a 20%, apenas 26,89% das ocorrências de leituras *on* teriam pistas fonéticas úteis—ou **22,43%** das ocorrências de *kanji* na amostra.

Estes são resultados preliminares, mas não favorecem a conjectura de DeFrancis sobre a escrita japonesa: os caracteres chineses parecem ser *menos* fonográficos no seu emprego japonês, não mais.<sup>19</sup> Os resultados são consistentes com as conclusões de Kess e Miyamoto (1999), que, após revisão da literatura psicolinguística, inferem que os dados (p. 200)

certamente sugerem diferenças nos estágios iniciais do reconhecimento de palavras escritas em *kanji* japoneses ou no alfabeto inglês. Enquanto a ativação fonológica acontece cedo e de forma eficiente na leitura de palavras inglesas, o mesmo pode não ser verdade para palavras japonesas em *kanji*. A ativação fonológica pode acontecer também nos *kanji*, mas não acontece a tempo de ativar sozinha a semântica, ou então a informação fonológica disponível é insuficiente para tal.

A despeito da presença de elementos não-fonológicos no alfabeto e de elementos fonológicos no *kanji*, os autores inferem ainda que (pp. 194–)

nós de forma postularíamos que o processamento mental do *kanji* é **essencialmente** o mesmo que o do alfabeto ou *kana*, especialmente nos estágios iniciais. Nossa revisão da literatura parece deixar óbvio que os processos de reconhecimento de padrões são frequentemente diferentes para estímulos do tipo logográfico [=nosso “morfográfico”], tipicamente apresentando maior dependência dos vários componentes visuais gráficos. [...] Dependendo do contexto de cada *kanji*, de sua familiaridade, frequência e complexidade, uma de duas rotas pode ser tomada como o caminho mais eficiente para o reconhecimento da palavra. Ambas as rotas terminam por recuperar informação semântica, sendo que uma é mediada pela fonologia e outra, pela informação grafológica. [grifo nosso].

A análise acima é um trabalho em andamento. É possível aperfeiçoá-la em diversos pontos, incluindo:

**Qualidade da listagem de componentes por *kanji*.** Usamos a base de dados do projeto KANJI VG,<sup>20</sup> mas ela ainda possui alguns erros quanto à divisão em compo-

<sup>19</sup> É claro que, em outro sentido, a escrita japonesa como um todo apresenta maior fonografia que a chinesa, por ser mista ou digráfica, empregando *kanji* e *kana*. Uma vez que os caracteres *kana* são especializados para fonografia, talvez seja natural que, por divisão de papéis, os *kanji* tenham se especializado para a morfografia.

<sup>20</sup> Apel et al. (2013).

centes. A equipe do projeto gentilmente aceitou nossas sugestões, mas é preciso uma revisão extensiva. Outro detalhe é que o critério do que conta como variação de um componente no KANJI VG nem sempre corresponde a nossos propósitos; por exemplo, em alguns casos, são usados critérios etimológicos não-óbvios para o leitor moderno.

**Comparação com outros sistemas de escrita.** Seria interessante medir a consistência fonética dos caracteres *kana*, e também da escrita portuguesa ou inglesa, e comparar os resultados com os *kanji* japoneses. Uma diferença grande nos índices pode reforçar o argumento que há distinção categórica.

**Leituras Jōyō especiais.** A tabela oficial do Jōyō Kanji <sup>21</sup> está em um formato difícil de tratar por máquina, particularmente a coluna “notas” (参考). Apesar do nosso esforço com o JoyoDB, um número de leituras excepcionais ficou excluída de nossos modelos. Além disso, nem sempre a lematização do analisador morfológico (Mecab+Unidic) concordou com a forma escolhida como lema pela tabela Jōyō. Isso tudo significa que nossa análise subestima o número de leituras Jōyō. Felizmente a proporção não-analisável foi baixa, e, além disso, esses problemas dizem respeito a leituras *kun* e especiais, que não são o foco desta análise em particular. Seria recomendável, de toda forma, ampliar o alcance do JoyoDB.

**Componentes fonéticos irregulares.** Muitos componentes fonéticos sugerem uma leitura aproximada, inexata. A contagem original de DeFrancis considerou como pista fonética qualquer leitura com pelo menos metade dos fonemas da leitura desejada. Outros autores definem critérios baseados na teoria de silabação da China histórica: inicial (ataque), final (rima+coda), e tom.<sup>22</sup> Não está claro como quantificar pistas irregulares no japonês, mas seria interessante propor um índice de regularidade e relacioná-lo à consistência. Esta melhoria tem o potencial de aumentar o grau de foneticidade da análise.

**Análise multivariada.** Uma análise estatística mais sofisticada, inferencial, poderia incluir nossos pressupostos como variáveis: ao invés de assumir que o leitor sabe qual o componente fonético, tentar modelar o fato que os componentes fonéticos tendem a estar em determinadas posições, incluindo uma variável “posição do componente”, e variáveis semelhantes para todos os pressupostos. Este tipo de análise provavelmente encontrará menor grau de foneticidade. Seria interessante relacionar dados como o tamanho da série fonética, a consistência, o grau de regularidade, o número de sílabas (ou moras), e alguma medida de dispersão categorial.

O código e os dados utilizados nesta análise estão disponibilizados no sítio desta dissertação.

---

<sup>21</sup> MEXT (2010)

<sup>22</sup> Chen et al. (2003).



## 5 Técnicas de expressão criativa usando *kanji*

Vimos na seção 3.5 (página 104) que o sistema de escrita japonês moderno combina dois tipos de caracteres, *kana* e *kanji*, que representam a língua em dois níveis, o fonográfico e morfográfico. A tabela abaixo revê as principais formas de uso do *kanji* no japonês moderno:

**Tabela 51 – Formas de uso regulares do *kanji* na escrita japonesa moderna**

Fonografia	Morfografia		Logografia/Semasiografia
	Japonesa	Sinítica	
<i>Ateji</i> “Caracteres atribuídos”	<i>On’yomi</i> “Leitura pelo som”	<i>Kun’yomi</i> “Leitura explicativa, tradução”	<i>Jukujikun</i> “Leitura convencionalizada”

Estas forma de uso são regulares; ou seja, são prescritas por documentos governamentais como o Jōyō Kanji (MEXT, 2010), ensinadas no sistema educativo, consagradas em dicionários e guias de estilo, etc. Mas, além dessas formas, a riqueza de combinações possíveis na escrita japonesa leva naturalmente à exploração criativa de outros modos de expressão. Há exemplos de *kanji* que fazem trocadilhos matemáticos, ou que desafiam a divisão entre fonograma e morfograma; argumentos são construídos com base nos componentes de um caractere, e ocasionalmente um *kanji* é usado como um autêntico símbolo extralinguístico, um ideograma.

Nesta seção, apresentamos uma classificação de algumas formas de expressão que, através do *kanji*, extrapolam o uso regular e criam vários efeitos estéticos ou subjetivos.

### 5.1 Reanálise morfológica

Na linguística, denomina-se “reanálise” o processo pelo qual os falantes alteram a análise morfológica de uma palavra, sem mudar sua forma sonora.<sup>1</sup> Ou seja, a palavra continua a mesma, mas um outro conjunto de morfemas é proposto para sua composição. Por exemplo, a tradição cristã propôs que a palavra “religião” (*religiō*, em latim) decompor-se-ia em *re-ligāre*, “re-ligar”, conceitualizando a religiosidade como a recuperação de uma conexão perdida com o divino. Lactâncio explicitamente rejeitou a análise proposta anteriormente por Cícero:

<sup>1</sup> Ver e.g. Haspelmath (1994).

*Hoc uinculo pietatis obstricti deo, & religati sumus. Unde ipsa religio nomen accepit, nō ut Cicero interpretatus est, a relegēdo, qui in libro De Natura Deorum secūdo dicitur.*

A religião é este vínculo da piedade que prende a Deus, religando-nos. Donde advém o nome “religião”—e não de *relegēdo* “reler”, como quer Cícero em seu livro *Sobre a natureza dos deuses*.

— *Diuinarum Institutionum Libri VII*, livro *iv*, capítulo 28 (SCHOLTEN; GÍSLADÓTTIR, 2015).

Todavia, pesquisas modernas mostram que Cícero estava correto sobre a formação da palavra; *re-ligiō* “religião” é originalmente o mesmo que *re-legēdo* “reler, ler repetidas vezes”, ambas derivadas da raiz indo-européia *h<sub>2</sub>leg* (op.cit.). Esse tipo de argumento milenar mostra que a separação de uma palavra em morfemas nem sempre é óbvia, e os falantes podem subentender análises alternativas.

Para um exemplo moderno, considere a palavra brasileira “bebemorar”, construída ludicamente a partir de uma reanálise de “comemorar”: originalmente formada por “com memória”, nesta brincadeira os fonemas iniciais /kome/ foram reanalisados como parte do morfema “come(r)”, que pôde então ser substituído por “bebe(r)”.<sup>2</sup>

Qual a relevância disto para o *kanji*? Acontece que, em uma escrita fonográfica, a reanálise é invisível: não há como distinguir na ortografia portuguesa o “comemorar” = “com memória” do “comemorar” = “comer”.<sup>3</sup> A morfografia, em contraste, representa uma língua no nível dos morfemas, o que abre a possibilidade de tornar as reanálises explícitas—e de usá-las para fins expressivos, como veremos a seguir.

### 5.1.1 Reanálise semântica

Chamaremos de “semântica” àquela reanálise que propõe um novo sentido, e este sentido é de alguma forma apropriado para aquele contexto. Por exemplo, considere o título do romance *Hotaru no Haka*, “Cemitério dos Vagalumes” (NOSAKA, 1967; TAKAHATA, 1988):

<sup>2</sup> Mais exemplos deste tipo são dados em Gonçalves, Andrade e Almeida (2010).

<sup>3</sup> Uma possibilidade seria o uso de hífen: “come-morar”. Porém, além de não-ortográfico, este sinal de pontuação não corresponde a nada na fala; ou seja, em todo caso, é necessário extrapolar o princípio fonográfico para grafar a reanálise. No caso da escrita morfográfica, essa possibilidade está sempre disponível, pois a grafia não-fonográfica faz parte do sistema. (Agradeço a Paulo Chagas pela observação.)

Figura 12 – Título da obra *Hotaru no Haka* (“Cemitério dos Vaga-lumes”)

火垂る の 墓  
*Ho-taru no Haka*  
 fogo-pinga GEN cemitério  
 “Cemitério dos Vaga-lumes”

*Hotaru no Haka* significa “cemitério dos vaga-lumes”. A trama do romance, trágica, gira em torno dos bombardeios incendiários de Tóquio durante Segunda Guerra. Os “vaga-lumes” do título aludem à crença de que vagalumes são as almas dos que morreram pelo fogo. “Vaga-lume” em japonês é *hotaru*, palavra normalmente grafada com um único *kanji*, 蛍—o que implica uma análise como morfema singular e indivisível. O livro em questão, porém, grafa-a como 火垂る *ho-ta.ru*, com caracteres que representam os morfemas “fogo” e “gotejar”; sem mudar o som, o autor faz com que “vaga-lume” seja lido como “pingos de fogo”, remetendo à violência das bombas.

Esse tipo de operação não é exclusiva da literatura. Considere o videogame *Shin Megami Tensei IV*, de público-alvo adolescente e temática de fantasia e terror (YAMAI, 2013). Neste jogo, os jogadores batalham contra monstros, mas podem também negociar com as criaturas e formar pactos. Observe a primeira palavra do seguinte anúncio comercial:

Figura 13 – Trecho de anúncio publicitário do videogame *Shin Megami Tensei IV*

仲魔            と 共 に  
*Naka-ma*        *to tomo ni*  
 Íntimo-demônios COM junto ADV  
 “Junto de companheiros”

強大            な 敵 に  
*Kyō-dai*        *na Teki ni*  
 forte-grande COP inimigos OBJ  
 “poderosos inimigos”

立ち            向かえ  
*tach-i*            *muka-e*  
 levantar-CON encarar-IMP  
 “encare!”

A palavra japonesa para “companheiro” é *nakama*, normalmente escrita 仲間; os *kanji* representam os morfemas *naka* “relacionamento” e *ma* “intervalo, espaço”. Isto é, a própria ortografia pressupõe e induz à análise de *nakama* como “dentro do espaço dos relacionamentos pessoais”.<sup>4</sup> Porém, no videogame em questão a palavra é grafada di-

<sup>4</sup> A idéia de “dentro” é reforçada visualmente pela composição do caractere 仲, que juxtapõe “pessoa” 亻 a “dentro” 中; e também pela homofonia de *naka* com o morfema “dentro”, grafado 中. De fato, a especialização heterográfica de 仲 *naka* = “relacionamento” como distinto de 中 *naka* = “dentro” deve ser posterior; etimologicamente *naka-ma* deve ser simplesmente “[alguém] dentro do espaço [pessoal]”.

ferentemente, como 仲魔. O caractere 魔 representa o morfema *ma*, “demônio”; *nakama* é então reanalisada como “relacionamento-demoníaco, monstro compactuado”—tudo sem alterar o som.

A presença abundante deste tipo de técnica no entretenimento infanto-juvenil—jogos, quadrinhos, *light novels*—é um dado importante, pois indica que tais formas de expressão com *kanji* não se restringem a sofisticadas manobras poéticas para eruditos. A ilustração acima, por exemplo, é de um anúncio de revista do jogo *Shin Megami Tensei*; isso implica que os publicitários confiam que seu público adolescente (ainda em processo de letramento de *kanji*) entenderá sem dificuldades os jogos de sentido. Por este valor para a análise, citaremos vários exemplos de mídia infanto-juvenil.

Para mais um estudo de caso, também na área publicitária, considere o seguinte anúncio da história em quadrinhos *Ansatsu Kyōshitsu* (“A Classe dos Assassinos”) (SHŪEISHA, 2012):

Figura 14 – Reanálise morfológica em *Ansatsu Kyōshitsu*

ジャンプ  
 J で 奇妙 な 授業 が  
 Janpu de Kimyō na Jugyō ga  
 Jump LOC estranha COP aula NOM

始まりそう な 今日 この 殺。。。  
 hajimari-sō na Kyō kono Koro...  
 começar-DEDU COP hoje esta hora

“Parece que está na hora de começar uma estranha aula na [nossa revista] *Jump*...”

O ponto relevante para nossa discussão é a locução adverbial de tempo no final do enunciado, *kono koro* “neste momento, nesta hora, por volta de agora”. Compare as palavras a seguir (com as glosas romanizadas):

Tabela 52 – Grafia das palavras *koro* “momento, hora” e *korosu* “assassinar”

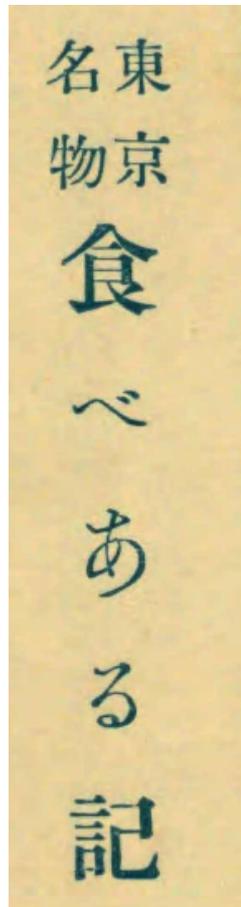
<i>koro</i> “momento, hora”	koro 頃
<i>koros-u</i> “assassinar”	koro su 殺す

A sequência fonológica *koro* é comum à palavra “momento, hora” e à raiz do verbo “assassinar”. Morfograficamente, a primeira é escrita com o caractere 頃 e a se-

gunda, com 殺.<sup>5</sup> No texto do anúncio, a expressão “nesta hora” troca o *kanji* de “hora” pelo de “assassinato”. Lido em voz alta, o efeito é imperceptível (até porque a raiz *koro-* de “assassinar” não é um morfema livre, e normalmente não poderia aparecer nessa posição em linguagem natural). Mas graças à sugestão do *kanji*, o texto implica uma camada de sentido extra: ao invés de “nesta hora de hoje”, há uma sugestão silenciosa de “neste assassinato de hoje”.

A técnica de reanálise não é província exclusiva da linguagem artística contemporânea. Considere o título do seguinte guia de restaurantes do período pré-guerra (JIJI SHINBŌ, 1929):<sup>6</sup>

Figura 15 – Capa do guia de restaurantes *Tōkyō Meibutsu Tabearuki* 東京名物食べある記, de 1929



A sequência fonética *tabearuki* normalmente seria analisada da seguinte forma:

<sup>5</sup> Note que, embora a divisão morfológica do verbo “assassinar” seja *koros-u*, devido a limitações do *kana* (que não pode representar consoantes independentes), a divisão ortográfica é *koro.su*. Assim, a escrita acaba criando uma pseudo-raiz *koro-*, distinguida graficamente de um pseudo-sufixo *-su* pela mudança de *kanji* para *kana*.

<sup>6</sup> Agradeço a Victor Mair, Nathan Hopson e ao usuário “V” do sítio *LanguageLog* por este exemplo.

## (5.1) 食べ 歩き

*tabe- aruk-i*

comer andar-NOM

“Passeio gastronômico”

Contudo, a ausência do morfograma esperado 歩 *aruk-* “caminhar”, somada à presença do morfograma 記 *ki* “documento, registro, diário”, transmite aos leitores uma análise alternativa:

## (5.2) 食べ ある 記

*tabe- aru ki*

comer EXP diário

“Diário de experiências gastronômicas”

## 5.1.2 Reanálise lúdica

A reanálise morfológica também pode ser semanticamente absurda, *nonsense*, fazendo como um jogo entre autor e leitor. O escritor Natsume Sōseki ficou conhecido por usar formas como *tonikaku* “de todo modo” = 兎に角 *to-ni-kaku* “chifre no coelho”, ou *yakamashii* “barulhento” = 八釜しい *ya-kama-shii* “com oito chaleiras” (LOOKVISE, 2016). Esse tipo de humor é muito presente na obra de Sōseki como um todo, não só no nível da escrita; considere, por exemplo, quando ele comenta o conto *La parule*, de Maupassant, e descreve a protagonista, uma senhora francesa, como saindo “com uma cesta de bambu para comprar tofu ocidental” (TREYVAULD, 2015). Esta “tradução”, é claro, não se refere a nada que efetivamente exista no texto original; trata-se de uma brincadeira absurdista, incongruente, que cria um elo de cumplicidade com o leitor. A reanálise de palavras com *kanji* criativos é congruente a esse estilo, ajudando a manter o mesmo tipo de distanciamento desconstrutivista do texto.

O humor também é popular em linguagem de Internet. Por exemplo, um marcador de “risos” em mensagens eletrônicas é (笑)—usando o *kanji* correspondente à palavra *warai*, “risos”. Acontece que normalmente esta palavra seria escrita como 笑い, com a flexão nominalizadora *-i* obrigatoriamente manifestada em *kana*. Por isso, quando digitado sem o *-i*, o *software* de conversão para *kanji* do computador não busca 笑 para *wara-*, sugerindo ao invés disso a palavra livre *wara* “palha”, escrita como 藁. Ou seja, ao tentar escrever 笑 *wara-* “riso”, o computador sugere primeiro 藁 *wara* “palha”. Como resultado, surgiu na cultura *online* a brincadeira de indicar “risos” pela notação (藁): “(palha)” (R25, 2010). Similarmente, no famoso fórum 2channel, a área para postagens de desenhos bidimensionais—em japonês, 二次元 *ni-jigen*—é com frequência abreviada como 虹 *niji* “arco-íris”, palavra homófona e incongruente, tanto por facilidade de digitação quanto pelo sentido de humor no absurdo (NICONICO, 2015).

### 5.1.3 Reanálise de estrangeirismo

É possível usar *kanji* para reanalisar um estrangeirismo, como em uma etimologia popular deliberada. Um exemplo favorito por Sōseki é *rōman*, de “romance”, escrito como 浪漫 = “vagas profusas”. A banda de *heavy metal* Onmyō-za, cuja temática é o Japão antigo, lança seus álbuns ao vivo (*live*, japonizado como *raibu*) sob a grafia 雷舞 *rai-bu* = “dança do trovão” (ONMYŌZA, 2015). O escritor Edogawa Ranpo adotou um nome que à primeira vista soa bem japonês, efeito reforçado pela ortografia tradicional em *kanji*: 江戸川乱歩, literalmente “passeio às margens do rio de Edo”. Edogawa era autor de contos policiais, e é preciso um olhar de detetive para enxergar que seu pseudônimo é uma homenagem cifrada: *Edogā Aran Po* = Edgar Allan Poe.

## 5.2 Paralelismo

Vimos que a escrita japonesa permite acrescentar glosas fonográficas, ou *furigana*, para especificar as leituras dos *kanji*—ou seja, quais morfemas estão sendo representados, e como pronunciá-los (sec. 3.5.7 à página 123). Até agora, só analisamos as glosas regulares, nas quais os *furigana* se restringem às leituras dicionarizadas dos seus *kanji*. Porém, o sistema de *furigana* também abre a possibilidade de inovar nas leituras atribuídas. É o que veremos a seguir.

### 5.2.1 Com palavras japonesas

Na obra de ficção *Nansō Satomi Hakkenden*, de Bakin (século XIX), aparece a seguinte frase (ARIGA, 1989):<sup>7</sup>

そ	は	又	えう 要	なき	おくりもの 人情
So	wa	mata	yō	nak-i	okur-i-mono[nin-jō]

Essa TOP ainda necessário não-ADN presentear-CON-coisa[humano-sentimento]  
 “Trata-se de uma recompensa[benevolência] desnecessária.”

O sentido superficial é “Trata-se de uma recompensa desnecessária”. Um guerreiro samurai está agradecendo ao senhor pela recompensa que recebeu no cumprimento do dever. Porém, a palavra *okurimono* “presente, recompensa” aparece como leitura *furigana* para os *kanji* 人情—*kanji* que não representam os morfemas de *okurimono*, como seria esperado, mas sim os morfemas *nin* “humano” e *jō* sentimento, formando a palavra *Ninjō* “Benevolência” (uma das Cinco Virtudes confucianas). Lembre-se da

<sup>7</sup> A grafia da glosa *furigana* transcrita aqui como *yō* é literalmente *eu*, refletindo a pronúncia do período Heian (ca. século IX). Porém, uma mudança fonética transformou o som em *yō*, por volta do século XIV, e é essa pronúncia que transcrevemos. (FRELLESVIG, 2010, p. 414–415).

divisão de trabalho da escrita japonesa com glosas: *furigana* para som e *kanji* para sentido. Como a posição *furigana* é especializada para a leitura fonográfica em *kana*, i.e. para o nível fonológico, entendemos que o samurai pronunciou a palavra *okurimono*. Mas como os *kanji* abaixo de uma glosa (na “base” da glosa) especificam quais são os morfemas—ou seja, o sentido—daqueles fonemas, o leitor entende que o sentido subjacente, implícito, não-pronunciado da “recompensa” é “benevolência”: o samurai não está emocionado apenas com o presente material em si, mas sim com a demonstração de generosidade do senhor.

Esta técnica também é usada na cultura popular moderna. Por exemplo, no volume 6 do mangá infanto-juvenil *Rurōni Kenshin*, aparece a seguinte fala (WATSUKI, 1995):

Figura 16 – Paralelismo no mangá *Rurōni Kenshin*



おとな                      ここ                      に                      いろ!  
 大人しく                      東京                      東京                      いろ!  
 Otonashi-ku                      koko[Tōkyō]                      ni                      ir-o!  
 Comportado-ADV                      aqui[Tóquio]                      LOC                      ficar-IMP  
 “Comporte-se e permaneça aqui [=em Tóquio]!”

A palavra “aqui” está grafada como a leitura *furigana* *ここ koko*, que acompanha irregularmente os *kanji* 東京, “Tóquio”. O contexto é que uma guerra está explodindo em Kyoto, e a personagem está ordenando rudemente a outra que não vá até o campo de batalha, pois ela seria inútil no combate. O que sai da boca da personagem é “aqui”, mas os *kanji* nos dão o sentido subjacente ao dêitico: fique “aqui” = em Tóquio, longe do campo de batalha.

Representamos o paralelismo com colchetes nas traduções, mas colchetes são lidos sequencialmente. No caso do *furigana*, a sensação subjetiva é de simultaneidade: ambas as palavras ocorrem ao mesmo tempo na mesma posição sintática, algo impossível na fala—o aparelho vocal humano só pode produzir um som de cada vez, e portanto a língua falada é necessariamente linear. O efeito estético de simultaneidade deve-se ao fato do leitor ser treinado para encontrar os fonemas nas glosas, e os morfemas correspondentes nos *kanjis*; ou seja, depende do princípio morfográfico. Pode-se especular que o efeito esteja ligado, talvez, à superposição das duas rotas neurais de leitura: a rota lexical e a fonológica (NAKAMURA et al., 2005).

O paralelismo na escrita japonesa é, portanto, uma forma de expressão que, por um lado, é inegavelmente linguística (uma superposição de dois signos na mesma posição sintática); mas, ao mesmo tempo, é intrínseca à escrita, sem nenhum correspondente na fala. Ou seja, a escrita aqui está representando a língua, mas não a fala. Embora a fala sonora não tenha o potencial para paralelismo, uma analogia intrigante pode ser feita com as línguas de sinais—que, por disporem de mais de um articulador independente, frequentemente fazem uso de elementos simultâneos; por exemplo, muitas línguas de sinais permitem a negação de um morfema através da coarticulação de um gesto manual com um menear de cabeça (ZESHAN, 2004). Por isso, ABNER argumenta que a linearidade da fala não é uma propriedade intrínseca à língua, mas simplesmente uma limitação accidental do meio de expressão sonoro. Sendo assim, é natural que a escrita, livre de tal limitação, faça uso da possibilidade da expressão simultânea, bidimensional. Podemos notar ainda que tanto as línguas de sinais quanto a escrita em *kanji* são, se comparadas à fala, bastante mais icônicas, não arbitrarias; isto é, empregam muitos significantes motivados, cuja forma remete ao significado (sejam os pictogramas usados como componentes de *kanji*, sejam os gestos imitativos comuns em línguas sinalizadas). Podemos supor que os significantes icônicos são pouco comuns na fala simplesmente porque é difícil sustentar muita iconicidade através do meio sonoro (mas cf. nota sobre simbolismo sonoro à página 33).

### 5.2.2 Com estrangeirismos

Uma outra maneira de usar paralelismo com fins expressivos é combinar *kanji* com termos estrangeiros. Por exemplo, o conto de Haruki Murakami, *Sidney no Green Street* (2006), começa com o seguinte trecho (ênfase e tradução nossas):

シドニーのグリーン・ストリートはあなたがそ  
の名前から想像するほど—たぶんそう想像するんじゃ

ないかと僕は想像するわけなのだけれど—素敵な通り  
 じゃない。だいいちこの通りには木なんてただの一本  
 も生えちゃいない。芝生も公園も水飲み場もない。な  
 グリーン・ストリート  
 のにどうして「緑通り」などというたいそうな名前が  
 付くことになったのか、これはもう神様でもなくちゃ  
 わからない。神様だってわからないかもしれない。

A *Green Street* de Sidney não é tão bela quanto o nome faz imaginar (ou pelo menos quanto eu imagino que o nome lhe fará imaginar). Em primeiro lugar, não há nesta rua uma única árvore sequer. Tampouco há gramados, parques, ou fontes d'água. Mas então, com tudo isso, por que será que a rua ganhou um nome tão exuberante quanto *Green Street* [Rua Verde]? O motivo, só Deus sabe. Quiçá nem mesmo Deus saiba.

Considere a primeira menção do nome da rua:

- (5.3) シドニー の グリーン ・ ストリート は [...]  
*Shidonii no Guriin · Sutoriito wa [...]*  
 Sidney GEN Green Street TOP  
 “A *Green Street* de Sidney [...]”

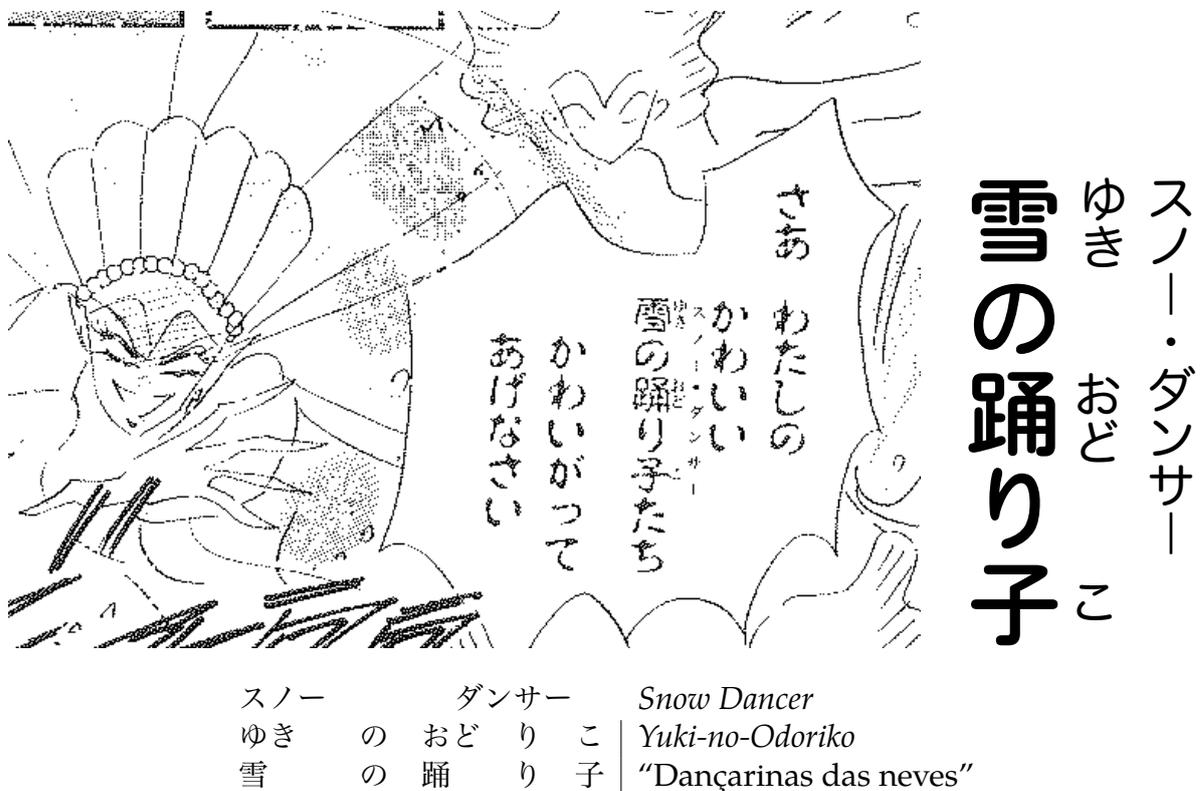
Todas as palavras estão grafadas fonemicamente em *kana*, com os nomes próprios estrangeiros transcritos em *katakana*, como é a norma: *Green Street* = *Guriin Sutoriito* = グリーン・ストリート. O narrador continua, descrevendo a rua como urbana e deserta e, algumas linhas depois, comenta:

(5.4)

なのに どうして <sup>グリーン・ストリート</sup> 『緑通り』 など と  
*Nanoni dōshite Guriin Sutoriito[midori-dōri] nado to*  
 Apesar disso por que Green Street[verde-rua] algo como quot  
 いう [...]  
*iu*  
 chamar  
 “Então por que será que, apesar de tudo, ela foi chamada de  
*Green Street* [Rua Verde][...]”

Desta vez, o nome estrangeiro Green Street não está grafado apenas em fonografia simples, como *kana*, mas aparece como glosa *furigana* para a palavra japonesa em *kanji* 緑通り *midori-doori*: “rua verde”. Sabemos que o narrador pronunciou Green Street porque o *furigana* indica a pronúncia; mas, ao contrário da primeira passagem, neste ponto o texto quer enfatizar o significado das palavras em inglês, e para isso ilustra-o em morfogramas *kanji*. Normalmente o *furigana* serve para ajudar a ler um *kanji* raro ou difícil, mas, na técnica do paralelismo com termo estrangeiro, as funções se invertem; é o *kanji* que ajuda a entender o sentido de um estrangeirismo, que o autor presume pouco familiar. Essa função explicativa do *kanji* fica clara na seguinte cena da história em quadrinhos *Saior Moon*, que usa dois níveis de *furigana* (TAKEUCHI, 1994):

Figura 17 – *Furigana* duplo com estrangeirismo em *Bishōjo Senshi Sailor Moon*



A autora usou a expressão *kanji* 雪の踊り子 como explicação para o nome *Snow Dancer*, transformado em glosa *katakana*. Contudo, ela deve ter julgado que sua audiência jovem talvez não tivesse o domínio de um *kanji* como 踊り *odori* “dança”, e por isso acrescentou um segundo nível de *furigana*, dessa vez no uso regular para indicar as leituras dos *kanji*.

Outra demonstração do efeito explicativo do *kanji*, quando usado como base para *furigana* estrangeiro, é a proposta feita por Hashimoto, Suzuki e Yamada (1987, p. 126), que sugerem a regularização de estrangeirismos em glosas como forma de facilitar sua compreensão. Alguns de seus exemplos incluem:

**Tabela 53 – Exemplos de estrangeirismos como glosa de *kanji* propostos por Hashimoto et al.**

Forma	Sentido dos <i>kanji</i>	Glosa/leitura	Origem	Significado
ハイテク 高技	alta-técnica	<i>hai-tekū</i>	<i>high-tech</i>	alta tecnologia
パソコン 個算	pessoal-calcular	<i>paso-kon</i>	<i>per(sonal) com(puter)</i>	PC
ワープロ 語処	palavra-processar	<i>wā-puro</i>	<i>wor(d) pro(cessor)</i>	processador de texto
ハイカラ 高襟	alto-gola	<i>hai-kara</i>	<i>high collar</i>	gola alta
ミスプリ 誤印	engano-imprimir	<i>misu-puri</i>	<i>missprint</i>	erro de impressão

Note que, nesta proposta, os estrangeirismos são escritos seguindo o princípio morfológico; e.g. o morfema *high* é sempre representado por 高 “alto”.

O esquema acima não encontrou uso amplo; porém, o processo de atribuição de estrangeirismos a *kanji* é perfeitamente paralelo ao processo *kanbun kundoku* que gerou as leituras *kun* (sec. 3.5.4) (donde os autores denominam sua proposta 昭訓 *Shōkun*, ou “leituras *kun* contemporâneas”). Assim, existe a possibilidade que algumas dessas associações tornem-se leituras *kun* dicionarizadas. De fato, o *kanji* 頁 “página” já consta como o morfograma de *pēji* (do inglês *page*), e certas outras associações têm se difundido bastante, como 超 “super, além” para *sūpā* (<*super*>) (NICONICO, 2016). Exemplos incluem o título da animação ドラゴンボール<sup>スーパー</sup>超 *Dragon Ball Super*,<sup>8</sup> um restaurante chamado<sup>スーパー</sup>ラーメン<sup>スーパー</sup>超とんこつ *Lamen Super Tonkotsu*,<sup>9</sup> ou um livro sobre técnicas de faxina, <sup>スーパー</sup>超おそうじ術 *Super Osōji-jutsu* “super técnicas de limpeza”:<sup>10</sup>

<sup>8</sup> Imagem: <<http://akkinews.net/archives/106245>>

<sup>9</sup> Imagem: <[http://izufull.com/shop\\_print.php?sid=r9W8hESx6z](http://izufull.com/shop_print.php?sid=r9W8hESx6z)>

<sup>10</sup> Imagem: <<https://www.kinokuniya.co.jp/f/dsg-01-9784391129878>>.

Figura 18 – O uso crescente de *super* スーパー como leitura do kanji 超: *Dragon Ball Super, Super Osōji-jutsu, Super Tonkotsu*



ドラゴンボール スーパー  
Dragon Ball 超



スーパー  
超おそうじ術



すーぱー  
超とんこつ

O autor Jun'ichirō Tanizaki valeu-se extensivamente do paralelismo com estrangeirismo em *Sasameyuki (As Irmãs Makioka)*, empregando-o não apenas em nível de palavra, mas em nível de enunciado (TANIZAKI, 1995, v. 中, ch 21, p. 169):

悦子は、

アインス ツワイドライ フィール

「一、二、三、四、.....」

などと、一から二三十ぐらいまでは独逸語で

数えを云い、

シュネル シュネル

「早く！早く！」

とか、

「ルミーさん、<sup>ビッテ</sup>何卒」

とか、

ノッホニヒト

「未だいけない！」

とか、そんな程度のことはその外にもいろいろ

ろ使った。

As falas acima são grafadas em japonês com *kanji*, mas sobre elas superpõe-se leituras em alemão na posição de *furigana*. Especificamente, vemos:

Tabela 54 – Uso de glosas em alemão em Sasameyuki: “ler com legendas”

	Base	Glosa
アインス 一	<i>Ichi</i> “um”	<i>Eins</i>
ツワイ 二 一	<i>Ni</i> “dois”	<i>Zwei</i>
ドライ 三 二	<i>San</i> “três”	<i>Drei</i>
フィール 四	<i>Shi</i> “quatro”	<i>Vier</i>
シュネル 早く	<i>Hayaku</i> “rápido”	<i>Schnell</i>
ビッテ 何卒	<i>Nani tozo</i> “por favor”	<i>Bitte</i>
ノッホニヒト 未だいけない	<i>Mada ikenai</i> “ainda não”	<i>Noch nicht</i>

A função do par *kanji/furigana* aqui é comparável ao uso de mecanismos como notas de rodapé, parênteses etc.; mas a sensação subjetiva, estética, é diferente. Compare com a tradução brasileira de Gotoda et al. (TANIZAKI, 2005, p. 356):

— *Eins, zwei, drei, vier...* — Etsuko contava em alemão de um a trinta.  
— *Schnell, schnell*, rápido — dizia. — *Bitte*, por favor, Rumi. *Noch nicht*, ainda não. — Além dessas, conseguia dizer mais algumas frases de mesmo nível.

As tradutoras combinam elegantemente o alemão, destacado em itálico, com a versão em português, interposta fluentemente através de vírgulas. Conquanto válida, a tradução necessariamente perde o efeito de paralelismo do original, que só é possível em uma escrita morfográfica com glosas. Através da técnica de paralelismo de Tanizaki, o leitor de japonês, acostumado a apreender simultaneamente a pronúncia via *furigana* e o sentido via *kanji*, tem a impressão de ler as duas línguas ao mesmo tempo. O efeito já foi descrito como “uma bizarra sensação de ler com legendas”.<sup>11</sup> Mesmo que a dis-

<sup>11</sup> Agradeço a David Eddyshaw pela expressão, e por apontar a técnica de Tanizaki (EDDYSHAW, 2013).

posição tipográfica fosse imitada em português, ela não surtiria efeito sem o hábito da leitura superposta de morfografia / fonografia, *kanji/kana*:

—<sup>Eins</sup>Um, <sup>zwei</sup>dois, <sup>drei</sup>três, <sup>vier</sup>quatro—Etsuko contava em alemão de um a trinta. —<sup>Schnell</sup>Rápido, <sup>schnell</sup>rápido—dizia. Por favor, Rumi. <sup>Noch nicht</sup>Ainda não.—Além dessas, conseguia dizer mais algumas frases de mesmo nível.

Novamente, a técnica tem presença na cultura popular tanto quanto na literária ou artística. Considere o exemplo da história em quadrinhos de fantasia, *Sailor Moon* (TAKEUCHI, 1994):

Figura 19 – Exemplo de paralelismo com língua estrangeira em história em quadrinhos: *Deep Submerge!*



A personagem invoca um ataque com nome em inglês, *Deep Submerge* (“submersão profunda”), grafado em *katakana* na posição de glosa. Isso permite adicionar um equivalente em *kanji*, 深水没 *Shin-suibotsu*, na mesma posição sintática.

Tampouco se restringem estes recursos às expressões artísticas ou de entretenimento; eles também se mostra úteis em outros tipos de linguagem, como a técnica ou acadêmica. Por exemplo, no livro de crítica literária *Koten Nihongo no Sekai* (OKAMOTO, 2007) aparecem os termos *National Literature* 国文学 e *National History* 国史:

想像が〈国民の言葉〉に及ぶのは半ば必然で

あり、同じことは「国文学」や

「国史」にも当てはまります。

Ambos são escritos na forma de leituras *furigana*, sobre as traduções em *kanji*: *kokubungaku* e *kokushi*. A combinação de fonogramas com morfogramas permite introduzir os termos técnicos (referentes ao estudo do nacionalismo do século XIX), e simultaneamente equipará-los com suas glosas.

### 5.2.3 Com outros tipos de caracteres

A forma padrão da glosa *furigana* ou *rubi* usa morfogramas *kanji* na posição base e fonogramas *kana* na posição de anotação. A forma é estável o suficiente que outros tipos de escrita aparecem em ambas as posições—sempre com a conotação de que a glosa indica a pronúncia (significante). Por exemplo, vimos no anúncio de *Ansatsu Kyōshitsu* (figura 14, p. 160) acima que o *furigana* foi usado para criar um trocadilho entre *koro* “momento” e *koro* “assassinato”. O mesmo anúncio inclui ainda um uso de glosa em *katakana* acoplada a uma base composta por uma letra latina (SHŪEISHA, 2012):

Figura 20 – Uso de abreviação em alfabeto romano com glosa *furigana*



ジャンプ  
 J で ...  
 Janpu de ...  
 Jump loc ...  
 “Na [nossa revista] *Jump* ...”

As fontes do alfabeto latino são projetadas para a leitura na horizontal, não vertical, e as formas típicas de cada letra são mais estreitas que os caracteres japoneses. Isso torna deslegante o *layout* de palavras alfabéticas inteiras no meio de texto vertical japonês. O uso de abreviações (inicialismos) permite usar uma única letra latina, em tamanho grande, como se fosse um *kanji*. Nesses casos o *furigana* serve para expandir a abreviação.

A história em quadrinhos *Green Blood* ilustra um exemplo de *furigana* invertido, no qual a posição base é ocupada por um nome próprio em *katakana* e a glosa é uma expressão em *kanji* (KAKIZAKI, 2011):

Figura 21 – Furigana invertido em *Green Blood*

ねえ、<sup>兄ちゃん</sup>ブラッド…  
 Nee, Buraddo[Nii-chan]…  
 Né, Brad[meu]  
 “Né, meu irmão [=Brad]”…

Como o *furigana* indica a pronúncia, sabemos que o enunciado falado foi a expressão de tratamento familiar *Nii-chan* 兄ちゃん (uma forma carinhosa de se dirigir a seu irmão mais velho). Porém, essa expressão está escrita como glosa de uma base em *katakana*: o nome do irmão, Brad. Isso é uma variante do exemplo em que glosas são usadas para resolver dêiticos; aqui, a notação serve como anotação, lembrando ao leitor de que o irmão do falante é a personagem chamada Brad, o que facilita acompanhar trama. Note como a posição base implica “sentido” (no caso, resolvendo a referência do pronome); isso reforça a análise de que o *kanji* é interpretado primariamente como grafia do sentido, a tal ponto que a sua posição normal no sistema de glosas passa a ser assumida como a posição que representa o sentido.

Um uso bastante criativo de *furigana* aparece no romance visual (*light novel*) *No Game No Life* (KAMIYA, 2012, p. 13):<sup>12</sup>

まことしやかに<sup>ささや</sup>囁かれる『<sup>くうはく</sup>』というゲームの噂だ。

Makoto-shiyaka ni sasayakareru “

”[Kuuhaku] to iu Gēmu no Uwasa da.

Pareciam ser reais os rumores sobre o jogo chamado “ ” [Vazio].

<sup>12</sup> Agradecimentos a @landorien <<https://kazemakase.ca/>> pelo exemplo.

O nome de um misterioso jogo online é apenas um espaço em branco; o *furigana* nos indica que isso se pronuncia *Kuuhaku*, “o vazio”.

### 5.3 Logografia e criatividade em nomes próprios

Nós analisamos a logografia *jukujikun* na seção 3.5.4.1 (p. 3.5.4.1). Nesta técnica, os *kanji* não representam morfemas individuais; eles são lidos em conjuntos que representam palavras inteiras (e.g. 大人, enquanto par indivisível, representa a palavra indivisível *otona* “adulto”; não é possível segmentar essa palavra em morfemas). Com frequência, a logografia japonesa lança mão de ideografia (ou semasiografia); ou seja, há uma relação de sentido entre os *kanji* e a palavra representada, mas não uma relação de pronúncia (e.g. 大 adiciona a idéia de “grande”, e 人, “pessoa”, para indicar “adulto”; mas nenhum dos dois *kanji* tem relação com a pronúncia *otona*). Pode-se talvez considerar a logografia como uma espécie de paralelismo sem glosas. É como se uma expressão como “pessoa grande”, *ookii hito* ou *\*daijin* (ambas leituras regulares de 大人), tivessem uma glosa de pronúncia especial *otona*.

Quando explorada de forma inovadora, a logografia em geral lançará mão de paralelismo com glosas *furigana*, ou o leitor não conheceria as novas leituras. Uma situação em que as glosas normalmente são omitidas são os nomes próprios, pessoais ou geográficos. Alguns nomes seguem grafias regulares, como o nome de família 森 Mori “bosque” ou a província de Hiroshima 広島 “Ilha Grande”. Outros usam leituras especiais, disponíveis apenas para nomes, padronizadas em documentos próprios;<sup>13</sup> nomes de família como 長谷川 Hasegawa ou 長谷見 Hasemi, e topônimos como 長谷寺 Hasedera ou 南長谷 Minamihase, sugerem as leituras *ha* e *se* para 長 e 谷; porém, essas são leituras ditas *nanori* 名乗り (“que vão em nomes”), que não ocorrem em outros contextos.

A pouca familiaridade das leituras *nanori* já seria suficiente para tornar a grafia dos nomes próprios confusa; porém, eles não se limitam a isso. Alguns nomes são logográficos, estilo *jukujikun*, sem relação sonora com os *kanji* individuais; e.g. o nome de família Hattori 服部 (*Fukube* “distrito das roupas”) ou a província de Kobe 神戸 (*Kamido* “portal dos deuses”). A associação entre *kanji* e leitura/morfema pode ser essencialmente arbitrária, necessitando decorar caso a caso. Recentemente, até nomes ocidentais tem sido introduzidos à grafia em *kanji*, como 空 “céu” = すかい Sky, 聖母 “mãe sagrada” = まりあ Maria, ou 海 “mar” = まりね Marine. Tais inovações tem sido criticadas como exageradas ou impróprias; porém, observando com mais atenção a história dos nomes japoneses, percebemos rapidamente que a criatividade com *kanji* tem sido a regra em toda geração (GROSS, 2015).

<sup>13</sup> Como a lista *Jinmeiyō Kanji*, “caracteres chineses para uso em nomes pessoais”, que regulamenta caracteres e leituras de uso exclusivo para nomes e sobrenomes.

Alguns nomes fazem operações com *kanji* que demonstram uma familiaridade profunda com a estrutura dos caracteres. Considere o nome da província de Tochigi 栃木, “árvore *tochi*”. Trata-se de um caractere criado no Japão (*kokuji*) para representar o nome da castanheira local, *Æsculus turbinata*. O caractere *tochi* é composto pelo componente semântico 木 “árvore”, adicionado a um uso incomum de 万 como fonético. Com o significado de “dez mil”, 万 normalmente indicaria a leitura *man* ou *ban*; mas aqui ela indica *tochi*, sob a lógica que “dez” *to* vezes “mil” *chi* = *tochi* = dez mil 万. Ou considere a grafia de ていな Tina como 茶奈, que usa *kanji* como fonogramas; o segundo, 奈, remonta à antiguidade como fonograma para o som *na*, mas o primeiro, “chá” 茶, normalmente é lido como “chá”; aqui ele usa a pronúncia inglesa, *tea* / ti: /, fazendo um rebus em cima de uma “leitura *kun* estrangeira”.<sup>14</sup>

## 5.4 Outras técnicas gráficas

As técnicas a seguir fazem uso da natureza gráfica da escrita. Elas incluem a escolha do tipo de caractere para criar efeitos de significado, a noção de peso e estilo.

### 5.4.1 O tipo de caractere como recurso expressivo

Em muitas obras de ficção, as falas de crianças são grafadas com *hiragana* onde normalmente se usaria *kanji*. Como uma criança em letramento não conhece muitos *kanji*, o uso excessivo de *hiragana* dá um ar infantil. Similarmente, estrangeiros com sotaque falam palavras japonesas em *katakana*, a grafia dos estrangeirismos—bem como vozes inumanas ou alienígenas, como robôs ou demônios. Em tais casos, a escolha do tipo de caractere tem valor semântico dentro do texto.

Um autor pode escrever um personagem repetindo uma palavra em *kana* quando ela não entendeu o significado; uma vez compreendida, ela passa a ser grafada em *kanji*. No conto *Kamekō-baka* (“A Tumba de Casco de Tartaruga”), do escritor okinawano Tatsuhiro Ōshiro, o senhor Eitarō, veterano militar, reverte ao jargão do Exército ao perceber que sua ilha fora invadida por aviões americanos; mas sua esposa Take não compreende a expressão que o marido usa (ŌSHIRO, 2002):

「誘導しているな」

「ユウドウては？」

<sup>14</sup> Esta operação é equivalente ao antigo fonograma rebus de leitura *kun*, o *kungana*, hoje em desuso; exceto que o *kun* em questão é um dos novos “*kun* estrangeiros”.

— *Yūdō-shiteiru na*

— *Yūdō te wa?*

— É um teleguiado.

— Tele o quê?

Em japonês, a palavra *yūdō* “guiar” (aqui abreviando *yūdōdan* “projétil guiado”) escreve-se morfograficamente como 誘導—como na fala de Eitarō, na primeira linha. Além de ser um linguajar estranho à esposa, o narrador nos conta que a palavra foi dita em japonês padrão (*hyōjungo*), não fazendo parte da língua local de Okinawa na qual conversavam até então. Por isso, Take repete a palavra sem entendê-la, interrogativamente; o autor torna isso claro ao usar, na sua fala (segunda linha), a grafia em *kana* ユドウ para a mesma palavra *yūdō*. Como o *kana* é fonográfico (isto é, baseado apenas no som e não no sentido), ele é mais apropriado para a situação na qual a personagem não sabe o sentido.

Recurso similar é empregado na obra *Genjitsu-sō* (“A Mansão Genjitsu”), de Masao Koike (KOIKE, 2011). A narradora está conversando por telefone com sua colega Yoriko, que acabou de dar a luz e está orgulhosa de ter feito parto natural:

通はテーオーセツカイといって、おなかを切るのよ。

よりこさんはそれだけは自慢していいことだというように、わたしに告げた。

ふうん、テーオーセツカイか、わたしはよりこさんほど丈夫じゃないから、産むとしたら、きっとそれね。頭のなかで、「テーオー」という音を、「帝王」という漢字に変換し、なぜ、こんなにっぽな名前が付いているのかしらといぶかしく思いながら、わたしはよりこさんにそう答えた。

“Normalmente eles fazem um negócio chamado *tēōsekkai* [テーオーセツカイ], em que cortam sua barriga, sabe?” Yoriko-san disse isso como se fosse algo de que pudesse se orgulhar pessoalmente.

“Hm, *tēōsekkai* [テーオーセツカイ]? Eu não sou assim tão forte quanto você, então no meu caso o parto provavelmente teria que ser desse jeito, né?” Enquanto respondia assim a Yoriko-san, **convertei [henkan] o som *tēō* na minha cabeça para os kanji 帝王 [“Imperador”]**, e fiquei pensando por que será que deram um nome tão imponente.

Quando a narradora ouve a palavra *tēōsekkai*, ela inicialmente não entende qual é o referente; isso é denotado pela grafia em *kana*, テーオーセツカイ, indicando que compreendeu o som, mas não alcançou o sentido. Depois, ela subitamente entende como é composta a palavra: *Tēō* “imperador” + *Sekkai* “cisão, corte” = “Corte Imperial” = cesariana. A palavra é então grafada em *kanji*, e além disso a personagem descreve a experiência linguística como “converter (*henkan*) o som para *kanji* na minha cabeça”. Isso sugere que, na concepção popular, o *kanji* é quintessencialmente uma grafia do sentido. Discutimos a concepção popular na seção 5.5 (p. 184) abaixo.

O recurso estilístico descrito aqui também não se limita à alta literatura. Considere esta cena do videogame *Ghost Trick* (TAKUMI, 2010), na qual a personagem conta para um fantasma qual era sua profissão em vida:

Figura 22 – Tipo de caractere como recurso expressivo: *Ghost Trick*



—そして…どうやら。

キミは“刑事”らしい。

—“けいじ”…って、あの、みんなのアカ  
ガレ、正義の見方の？

— Sendo assim... parece que você foi um “detetive” [刑事 *keiji*].

— “Detetive” [けいじ *keiji*]? Er, hmm, isso não era algo sobre ser um defensor da justiça, herói dos oprimidos, algo assim?

No começo da segunda fala, o fantasma não entende direito o que quer dizer a palavra *keiji* “detetive”. Isso é ilustrado pela maneira com que ecoa a palavra em *kana*, ao invés do *kanji* que aparece na fala de cima. Como visível pelo fluxo do discurso, é só imediatamente depois que o fantasma, que sofre de amnésia, se lembra do significado.

#### 5.4.2 Peso

A proporção de *kanji* em um texto está relacionada a sensações de formalidade, intelectualidade, seriedade, e de forma geral ao texto “pesado”. Além das consequências naturais para a retórica e o discurso, isso também é explorado por autores que podem, por exemplo, fazer uma personagem grave e séria falar com muitos caracteres *kanji*, e opô-la a uma personagem leve e casual que fale com muito *kana*.

Situações também podem estar refletidas no peso. Considere o poema de Anzai Fuyue (FUYUE, 1929):

てふてふが一匹韃靼海峡を渡って行った

てふてふ	が	一匹	韃靼	海峡	を	渡って	行った
<i>Chōchō</i>	<i>ga</i>	<i>ip-piki</i>	<i>Tattan</i>	<i>Kai-kyō</i>	<i>o</i>	<i>watat-te</i>	<i>it-ta</i>
Borboleta	NOM	um-CTR.animal	Tartária	mar-ravina	OBJ	CRUZAR-CON	IR-PRFV

“Uma borboleta cruzando o Estreito da Tartária e—foi.”

O verso começa com a palavra “borboleta”, *chōchō*, grafada em *kana* (na ortografia histórica, *tefu-tefu* < *tepu-tepu*; talvez por sugerir mimetismo com o movimento das asas). Quando a narrativa (e a borboleta) chega no trecho do Estreito da Tartária, a grafia muda para uma densa barreira de *kanji*, refletindo a travessia difícil; e, no final, quando ela escapa, os *kana* vão voltando, arejando a página com espaço branco. Este

tipo de técnica é comparável à poesia concreta brasileira, e de fato foi inspiração para ela (CAMPOS; DANTAS, 1994).

### 5.4.3 Estilo

É possível variar o tipo de caractere para provocar efeitos estilísticos. Uma alta proporção de estrangeirismos em *katakana* dá um ar de modernidade e sofisticação urbana cosmopolita, muito visado em publicidade contemporânea. Já uma alta proporção de *hiragana* gera uma impressão de leveza, podendo refletir casualidade ou infantilidade. Uma alta proporção de *kanji*, combinada com o emprego de caracteres raros ou históricos, traz um efeito classicista, pesado, erudito, literário; tal efeito é ampliado se for combinado com uma alta proporção de palavras siníticas (*kango*—isto é, leituras *on* dos *kanji*), que também dão um tom técnico, formal, intelectualizado. Se uma grande proporção de *kanji* é usada, mas incluindo muitas palavras de origem nativa (*yamato-kotoba*; leituras *kun*), o efeito continua literário e sofisticado, mas perde o peso do formalismo, do texto técnico, remetendo ao invés disso ao nacionalismo romântico, ao folclore e às tradições locais.

Um exemplo disso é o autor de romances policiais Natsuhiko Kyōgoku. Suas tramas fazem alusão às lendas de criaturas sobrenaturais japonesas, ou *yōkai*, cujos mitos são recriados na forma realista de criminosos, detetives e golpes elaborados. Kyōgoku é conhecido por empregar alta proporção de caracteres *kanji* raros, escritos com *software* especializado; para evitar que isso dificulte a leitura, o autor faz uso liberal de glosas *furigana*.<sup>15</sup>

Considere o exemplo abaixo, extraído do conto *Hakuzōsu*<sup>16</sup> (KYŌGOKU, 1999, p. 69):

<sup>15</sup> Exatamente a mesma combinação—muitos caracteres exóticos, mas sempre com glosas—foi lugar-comum durante a literatura popular pré-moderna, no período Edo; cf. Ariga (1989), Seeley (1991, pp. 100–101).

<sup>16</sup> *Hakuzōsu* 白蔵主 é um cargo dentro dos mosteiros Zen, cujo monge detentor é responsável pelo acervo de escrituras. O conto alude a uma famosa lenda sobre o Hakuzōsu do templo Kobayashi no século XIV, que cultuava a deusa Inari e cuidava de três raposas, animais sagrados desta divindade. As raposas terminam por revelar-se seres espirituais, e a história foi imortalizada na peça de teatro *kyōgen Tsurigitsune* 釣狐. A primeira frase do trecho de Kyōgoku é uma citação direta do texto original.

かい ゆめやま  
甲斐の国に夢山という名の山がある。

まこと もうろうもこ あお  
真に朦朧模糊として、仰ぐ者見る者

くらくら ひがん  
は、一様に夢夢と彼岸を感得し、分け

くらくら  
入る者歩く音は、ただ眩眩とするうち

なが くまじ いぎな  
に、生き乍ら隈路に誘われたが如き

くら  
心持ちになる。昼尚昏き闇こそない

そこここ うつしよ かくりよ とろ  
が、其処此処の現世と幽世の境が蕩けて

いて、故に夢山と呼ぶのである。

[Tradução livre] No feudo de Kai, há uma montanha conhecida como “Pico dos Sonhos”. Sua forma é verdadeiramente vaga e indistinta, de tal maneira que aqueles que a fitam do sopé ao topo são tomados por uma vertigem, e por um momento têm a impressão de ter vislumbrado a Outra Margem e contemplado o Nirvana; porém, ao ouvir o som dos passos de peregrinos que se embrenham montanha acima, percebe-se logo que a sensação vertiginosa nada mais era que uma tontura comum.

Esta pequena introdução de três frases usa 47 caracteres *kanji*, dos quais nada menos que 11 (cerca de 1/4) estão fora da lista Jōyō—e, portanto, não são ensinados na escola (ver tabela abaixo). Porém, todos estes estão dotados de glosas, de forma que o leitor tem fácil acesso ao texto mesmo que desconheça os caracteres. À primeira vista, o uso de formas do japonês clássico, como *-ga gotoki* (“semelhante a”), remete à literatura antiga; porém, um segundo olhar revela que se trata de japonês moderno arcaizado superficialmente. No nível do sistema de escrita, há glosas até mesmo para palavras regulares e relativamente comuns, como 仰ぐ *aogu* “olhar para cima” ou 彼岸 *Higan* “rito do equinócio; Nirvana” (ambas leituras listadas no Jōyō). Isso denuncia que a sensação de erudição é puramente estilística; a audiência do texto continua sendo a popular (especialidade de Kyōgoku).

É produtivo analisar caso a caso como o autor usa *kanji* e *furigana*:

**Tabela 55 – Técnicas de escrita de um trecho de *Hakuzōsu*, de Natsuhiko Kyōgoku.**

**Grafia**

**Análise**

かい

甲斐

Glosa para facilitar a leitura irregular: *Kai*. Nome de antiga província, em ortografia histórica, usando *kanji* como fonogramas (ou seja, desconsiderando o sentido dos caracteres, que seria “carapaça-padronada”). A leitura *ka* não é mais usada para o caractere 甲; mas a glosa *furigana* permite empregar essa leitura irregular sem trazer dificuldades ao leitor, e causando um efeito de romance histórico.

ゆめやま

夢山

Glosa regular: *Yume-yama* sonho-montanha = “Montanha dos Sonhos”. A posição presa dos caracteres (isto é, como parte de palavra composta) poderia sugerir leitura sinítica (*on*), o que daria *Mu-san*; mas a glosa clarifica que este topônimo se lê com pronúncia japonesa (*kun*). Isso reforça o efeito de regionalismo e folclore.

まこと

真に

Glosa regular: *Makoto* “verdadeiro”, seguido do adverbializador *ni*. Palavra de registro formal, mas bastante comum. A forma preferida pelo padrão *Jōyō* emprega outro caractere, 誠, de mesmo sentido; mas a ortografia usada por *Kyōgoku* é uma alternativa amplamente utilizada. A presença desta glosa sugere que o romance tem público-alvo popular; em um gênero literário considerado mais erudito, ela poderia ter sido omitida.

もうろう

朦朧

Glosa regular, originalmente um composto sinonímico: *Mōrō* enevoado-enevoado = “enevoado, obscuro, vago”. Palavra relativamente rara e caracteres difíceis, não-*Jōyō*, fazendo a glosa necessária.

もこ

模糊

Glosa regular: *Moko* = “indistinto, vago”. Esta palavra forma um composto sinonímico com a anterior: *Mōrō-moko* = “indistinto”. Sua etimologia morfológica, determinada pelos *kanji*, é pouco clara. O primeiro morfema, *mō* 模 “modelo”, alude à matriz de onde se produzem (por exemplo) xilogravuras; e por extensão, ao ato de produzir cópias. O segundo morfema, *rō* 糊, grafado com um caractere não-*Jōyō*, originalmente representa “cola”. Possivelmente a etimologia advenha da imagem de uma matriz (*mo*) preenchida por pasta (*ko*), de forma que suas cópias (*ko*) fiquem indistintas (*moko*).

くらくら  
夢夢

Técnica do paralelismo (cf. sec. 5.2): a glosa, *Kurakura*, não corresponde aos caracteres, que normalmente seriam lidos *Yume-yume*. Na superfície, a palavra pronunciada é *kurakura*, i.e. algo que dá tontura, vertigem; porém, a grafia do *kanji* subjacente corresponde ao adjetivo arcaico *yume-yume (shi)* “onírico, irreal”. A implicação é que a montanha causa vertigem a quem a olha de baixo para cima, não simplesmente por ser alta, mas por causar uma impressão de irrealidade ou sobrenaturalidade. Compare com a segunda grafia de *kurakura*, abaixo. Um detalhe tipográfico interessante é que a reduplicação está grafada por extenso—diferentemente do japonês moderno, onde caracteres *kanji* repetidos são abreviados pelo sinal 々. A grafia por extenso também reforça o efeito de classicismo.

くらくら  
眩眩

Desta vez, a palavra *kurakura* “vertiginoso” está grafada com o caractere 眩, que normalmente escreveria *kuramu* “desorientar-se”. Aqui, a sugestão onírica de *kurakura* = <sup>くらくら</sup>夢夢 é desmentida, e o autor volta explicitamente ao sentido prosaico de “algo que dá tontura”.

ながら  
生き乍

*Iki-nagara* “enquanto vivo, ainda nesta vida”. Em japonês moderno, elementos gramaticais como o sufixo *-nagara* (“enquanto”) não são grafados em *kanji*. Ao usar o caractere 乍 nesta função, *Kyōgoku* reforça o efeito arcaizante. Ao mesmo tempo, a glosa mantém a facilidade de leitura.

くまじ  
隈路

*Kumaji* curva-caminho = “caminho sinuoso, torto; caminho da iniqüidade”. Palavra incomum e potencialmente de difícil compreensão devido à homofonia (e.g. *kuma* poderia ser “urso”, *ji* poderia ser “terra” ou “hora” ou etc.); porém, a presença do *kanji* torna o sentido claro para o leitor, mesmo que não conheça o termo.

くら  
昏き

*Kuraki* “escuro”. O autor opta pelo caractere antigo não-Jōyō, 昏, ao invés da alternativa padrão moderna, 暗.

そこここ  
其処此処

*Soko-koko* “aqui e ali”. Outro exemplo de uso de *kanji* para dar mais classicismo e “peso” aonde, em língua moderna, se usaria apenas *kana* (em palavras gramaticais—no caso, os dêiticos de lugar).

うつしよ  
現世

*Utsushi-yo* “o mundo atual, a realidade mundana”. Termo budista baseado no adjetivo arcaico *utsushi* “real, verdadeiro”. Para um falante moderno, ele poderia facilmente confundir-se com homófonos como “cópia” (写) ou “transferência” (移し); porém, a grafia em *kanji* é idêntica ao equivalente moderno *Gense* 現世 “o mundo atual”, de forma que os caracteres novamente clarificam o sentido que o autor quer passar, permitindo o recurso ao arcaísmo sem causar confusão. Além disso, a forma arcaica permite um paralelismo fonético com *Kakuri-yo*, logo a seguir.

かくりよ  
幽世

*Kakuri-yo* “o outro mundo”. A etimologia desta palavra é “o mundo oculto”, escrita morfograficamente como 隱世. Porém, ao invés do caractere de *kakuri* “oculto”, 隱, o autor preferiu o caractere 幽, usado para grafar “espírito, fantasma”, ressaltando as conotações de sobrenaturalidade, além-vida, assombrações.

とろ  
蕩けて

*Torokete* “dissolve”. Mais um caractere não-Jōyō, cujo emprego é possível graças à glosa.

O ponto a tirar deste exemplo é como a escolha do tipo de caractere pode ser um recurso estilístico—no caso, para evocar uma atmosfera de classicismo etéreo, lirismo onírico e folclore regionalista; e, em geral, que as características complexas da escrita japonesa (*kanji*, escrita mista, glosas) são rotineiramente exploradas para criar efeitos expressivos.<sup>17</sup>

## 5.5 A escrita *kanji* na concepção popular

Baseado no tipo de técnica que vimos, convém repensar o que representa o *kanji* não apenas em modelos científicos, mas também na concepção popular em sua cultura nativa—e nas implicações dessas percepções subjetivas. Parece estar claro que, a despeito de reconhecer e empregar componentes fonéticos, esses caracteres não são sentidos pelos falantes como um tipo de fonografia.

Na seção 5.4.1 acima, sobre a escolha do tipo de caractere (*kanji* ou *kana*) como recurso expressivo, vimos um exemplo da obra *Genjitsu-sō* de Masao Koike. Além do uso de *katakana* para expressar incompreensão, vimos que a personagem usa expli-

<sup>17</sup> Kyōgoku é um autor de orientação visual, não só no que diz respeito ao sistema de escrita; ele é conhecido por controlar ele próprio, cuidadosamente, o *design* e *layout* de seus livros (KYŌGOKU, 2003). Por exemplo, para que o leitor não precise interromper a leitura, Kyōgoku insiste que nenhuma frase comece e termine em páginas diferentes.

tamente a expressão “converti para *kanji* na minha cabeça”, reproduzida aqui (KOIKE, 2011):

頭のなかで、「テーオー」という音を、「帝王」とい  
う漢字に変換し [...]

Converti [*henkan*] o som *tēō* na minha cabeça para os *kanji* 帝王 [“Imperador”] [...]

*Henkan* é a operação usada para digitar japonês no computador; as palavras são escritas fonologicamente, em *kana*, e quando o usuário inicia a operação de *henkan*, o computador converte para *kanji* morfológico, usando o contexto para ordenar as sugestões de grafias possíveis. A expressão é corrente no vernacular, e seu sentido é semelhante ao nosso “caiu a ficha”: o momento em que o significante alcança o significado. A noção está refletida também em outras expressões idiomáticas, como *henkan dekinakatta* “não consegui fazer *henkan*” = “não entendi o que isso quer dizer”, ou *ji mo shirazu* “não sabe nem o (*kan*)*ji* disso” = “está usando termos cujo sentido não entende”.

O seguinte trecho de SAITŌ (2014) (p. 95) torna mais explícito o raciocínio por trás da metáfora:

私たちは普段、意識していませんが、耳で聞いた言葉を頭の中で漢字仮名交じり変換することで理解しています。

テレビのニュース番組を、画面を見ないで耳だけで聞いていることはよくありますが、そういうときに、例えば、「ティーピーピーこうしょうがさいかいされました」と言われれば即座に頭の中で「T P P 交渉が再開されました」と変換しています。このとき「こうしょう」を「校章」や「考証」だと思う人はまずい

ません。でも、「交渉」という漢字を知らない子供は、この変換が行えず、自分の知っている言葉の中から当てはまるものを選ぶので、「こうしょうって、学校のバッジのこと？」と聞くかもしれません。

Ainda que normalmente isto se dê de forma inconsciente, o fato é que nós japoneses compreendemos as palavras que ouvimos através da conversão [変換 *henkan*], dentro de nossas cabeças, para uma mistura de *kanji* e *kana* [*kanji-kana majiri*; cf. sec. 3.5.6, pp. 122–].

Por exemplo, a maioria de nós já ouviu o noticiário da TV sem olhar para a tela. Se, em tal situação, a pessoa escutar o enunciado *Tiipiipii-Kōshō ga Saikai-saremashita* [em fonografia *kana* no original], ela instantaneamente o converterá dentro de sua cabeça para [a morfografia mista] T P P 交渉が再開されました “Negociações [*Kōshō*] sobre o TPP [*Tiipiipii*] foram reabertas [*Saikai*]”. Certamente ninguém pensaria em 校章 *Kōshō* “emblemata escolar” ou 考証 *Kōshō* “investigação”. Porém, uma criança que não conheça os caracteres *kanji* 交渉 *Kōshō* “negociações” não fará essa conversão *henkan*, mas escolherá uma palavra adequada dentre aquelas que conhece, e talvez pergunte algo como “Esse *Kōshō* [em *kana* no original] quer dizer o emblema da escola?”

Note como o autor mistura o conceito de “léxico mental”—as palavras e morfemas que indivíduo conhece—com “conhecer o *kanji*”, além de descrever a experiência de entender uma frase como “converter para texto morfográfico”. Do ponto de vista linguístico, essas imagens soam como a falácia de descrever a língua falada em termos escritos. Porém, dado resultados como os de Joyce (2004), a noção não parece tão absurda. Há evidência de que os grafemas dos *kanji* são armazenados mentalmente em grupos cognitivos (*clusters*) com os morfemas que representam, de tal forma que ouvir e compreender um morfema pré-ativa (*primes*) a representação gráfica. É possível, claro, conhecer um morfema sem saber seu *kanji* (cegos congênitos podem perfeitamente se tornar falantes hábeis de japonês); mas, no caso típico, uma criança japonesa vai aprender o significado de palavras eruditas como 交渉 *Kōshō* “negociação” junto com sua análise morfológica, ao ser ensinada sobre cada caractere. Ou seja, a metáfora “converter mentalmente para *kanji*” em muitos casos parece ser uma descrição aproximadamente válida—mais precisamente, o que acontece deve ser algo como “dividir a sequência de fonemas em subsequências, cada uma correspondendo a um agrupamento (*cluster*) de sons (incluindo outras pronúncias *on* e *kun*), sentidos, e caracteres”. O *kanji* é um símbolo perfeito desses agrupamentos cognitivos do léxico.

## 6 Conclusões

O modelo da transcrição da fala considera a escrita um signo de segunda ordem: uma representação indireta da língua falada. No decorrer deste trabalho, vimos inúmeras características das escritas que não podem ser reduzidas à transcrição da fala (como as estruturas sintático-semânticas específicas da pontuação, ou a heterografia como forma de acesso ao sentido). Tais recursos mostraram-se especialmente frequentes nos caracteres chineses, por serem baseados no nível morfológico. As especificidades da escrita refletem objetivos funcionais do meio, como a necessidade de clareza de sentido e facilidade de busca visual. Desta forma, o modelo da escrita como sistema autônomo de representação da linguagem, tal como defendido por [Vachek e Luelsdorff \(1989\)](#), [Halliday \(2014\)](#) ou [Nunberg \(1990\)](#), parece-nos explicar melhor os fatos.

Em particular, estudamos muitos exemplos de como os escritores da língua japonesa rotineiramente empregam formas de expressão que fazem uso do ferramental único à escrita morfográfica, e também técnicas de escrita mista (como o sistema de glosas *furigana*). Tais técnicas são por vezes tratadas como uma forma de comunicação elevada, própria da literatura ou da poesia clássica. Porém, vimos inúmeros exemplos de seu uso em cultura popular, mídia contemporânea e publicidade. O uso de recursos expressivos do *kanji* em anúncios e propaganda mostra que há expectativa de alta compreensão, e o uso ricamente documentado de tais técnicas em histórias em quadrinhos e videogames mostra que até mesmo aprendizes, ainda em processo de letramento, são considerados usuários sofisticados de *kanji*. Por tais fatores, nossa conclusão é que tais técnicas não são excepcionais, periféricas, ou arcanos jogos eruditos, mas sim parte integrante e vital do sistema de escrita japonês popular e cotidiano. Disso se segue que a escrita japonesa vai além da transcrição da fala.

Vimos que os exemplos mais extremos da crítica do *kanji* parecem acompanhar o comprometimento com o modelo da transcrição. Linguistas cujo paradigma enfatize a primazia da fala e o nativismo são levados naturalmente a adotar o modelo da transcrição, a fim de sustentar seus axiomas: se o paradigma teórico de um linguista leva-o a definir *a priori* que o papel da escrita é representar a fala, tida como única “língua verdadeira” e “natural”, ele só verá ineficiências desnecessárias em um sistema que faz mais do que representar a fala. Os adeptos de tais modelos parecem especialmente propensos a criticar os caracteres chineses, e particularmente a escrita japonesa, como algo ineficiente, custoso e desnecessário.

Para DeFrancis, 66% dos caracteres chineses são “fonéticos”—isto é, possuem um componente que indica sua leitura—o que, para o autor, é prova que o chinês (e,

por extensão, toda escrita possível) é fonográfico mais do que qualquer coisa. Ao tentar calcular um número aplicável ao japonês com métodos semelhantes ao do autor, e baseando-se em uma amostra real de tamanho amplo, encontramos apenas 22,43% de ocorrências de *kanji* com a leitura aproximadamente previsível, na melhor das hipóteses. Parece-nos claro que, conquanto o subsistema fonético seja real e importante, o mecanismo dos *kanji* continua sendo categoricamente distinto dos *kana*, provavelmente devido a diferenças de processamento neural tais quais as sugeridas por Nakamura et al. (2005), Handel (2015), ou Kess e Miyamoto (1999) (pp. 194–).

A escrita japonesa, mais até do que a chinesa, é abundante em formas de expressão que exploram a riqueza da escrita por todos os ângulos, combinando fonografia, morfografia, e por vezes até logografia e ideografia, aproveitando a bidimensionalidade do meio para sobrepor mais de um item linguístico na mesma posição sintática, causando todo tipo de efeito subjetivo. Tudo isso é sustentado por uma comunidade com grau muito alto de letramento, e por um grande e rico mercado editorial. É inegável que a aquisição da escrita *kanji* demanda mais tempo do que uma escrita fonográfica; porém, até o momento, a sociedade japonesa claramente tem demonstrado tanto a capacidade quanto a vontade de arcar com os custos do aprendizado. O resultado é, ao nosso ver, um tesouro cultural: uma comunidade letrada dotada de ferramentas que não se vê em nenhuma outra parte do mundo, formando um grande um laboratório de exploração da escrita. Matsunaga, discutindo a questão da abolição do *kanji*, argumenta que apenas “apegos sociais, culturais e emocionais” justificam a continuidade de seu uso (MATSUNAGA, 1994). Depois de analisar inúmeras formas de expressão que dependem dos *kanji*, e que seriam irrevogavelmente perdidas com uma reforma abolicionista, talvez esteja mais claro o que fundamenta tal sensação de “apego”. A experiência de ler um texto em escrita morfográfica mista é subjetivamente diferente da de lê-lo em fonografia pura; existem razões estruturais, neuronais e estéticas para tanto. Faz sentido, então, que uma cultura letrada seja sensível a essa diferença (um “fato subjetivo” inquestionável), e busque preservar o que sente ser algo valioso. A escrita em *kanji* é custosa, sim, mas custo não significa desperdício; ela também recebe algo em troca —algo que seria perdido em um mundo no qual ninguém mais escrevesse em *kanji*.

## Referências

- AARON, P. G.; JOSHI, R. M. Written language is as natural as spoken language: A biolinguistic perspective. *Reading Psychology*, London, v. 27, n. 4, p. 263–311, 2006. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/02702710600846803>>.
- ABNER, N. *Syntactic Categorization in Sign Languages*. Onex: Lingbuzz, 2016. Lingbuzz reference: lingbuzz/003189. Disponível em: <<http://ling.auf.net/lingbuzz/003189>>. Acesso em: julho de 2016.
- ACADEMIA BRASILEIRA DE LETRAS. *Vocabulário Ortográfico da Língua Portuguesa*. 5. ed. São Paulo: Global, 2009. Disponível em: <<http://www.academia.org.br/nossa-lingua/vocabulario-ortografico>>. Acesso em: dezembro de 2015.
- AL-RFOU, R. *Polyglot: Multilingual text (NLP) processing toolkit*. 2015. Acessado via pacote Pypi polyglot. Disponível em: <<http://polyglot-nlp.com>>. Acesso em: Julho de 2016.
- ALDRIDGE, E. *Analysis and Value of Hentai Kanbun as Japanese*. Washington: University of Washington Press, 2011.
- APEL, U. et al. *KanjiVG*. 2013. Disponível em: <<http://kanjivg.tagaini.net/>>.
- ARIGA, C. The playful gloss: Rubi in Japanese literature. *Monumenta Nipponica*, Sophia University, v. 44, n. 3, p. 309–335, 1989. ISSN 00270741. Disponível em: <<http://www.jstor.org/stable/2384611>>.
- ARISTÓTELES. *On Interpretation*. Tradução: E. M. Edghill. 2009. Disponível em: <<http://classics.mit.edu/Aristotle/interpretation.html>>.
- ARONOFF, M. Orthography and linguistic theory: The syntactic basis of Masoretic Hebrew punctuation. *Language*, Linguistic Society of America, Washington, v. 61, n. 1, p. pp. 28–72, 1985. ISSN 00978507. Disponível em: <<http://www.jstor.org/stable/413420>>.
- BACKHOUSE, A. E. Aspects of the graphological structure of Japanese. *Visible Language*, University of Cincinnati, Cincinnati, v. 18, n. 3, p. 219–228, 1984. Disponível em: <<http://visiblelanguagejournal.com/issues/issue/71/>>.
- BENTLEY, J. R. The origin of Man'yōgana. *Bulletin of the School of Oriental and African Studies, University of London*, Cambridge University Press on behalf of School of Oriental and African Studies, Cambridge, v. 64, n. 1, p. pp. 59–73, 2001. ISSN 0041977X. Disponível em: <<http://www.jstor.org/stable/3657541>>. Acesso em: julho de 2016.
- BERLIN, B. *Ethnobiological Classification*. Princeton: Princeton University, 1992.
- BLOOMFIELD, L. *Language*. New York: Holt, Rinehart and Winston, 1933.

- BOHM, A.; STATEN, H.; CHOW, R. Derrida and Chinese writing. *PMLA*, Modern Language Association, New York, v. 116, n. 3, p. pp. 657–660, 2001. ISSN 00308129. Disponível em: <<http://www.jstor.org/stable/463507>>.
- BOIKO, L. *yomisplit*. 2015. Disponível em: <<https://github.com/leoboiko/yomisplit>>. Acesso em: agosto de 2015.
- BOIKO, L. *JoyoDB*. 2016. Disponível em: <<https://github.com/leoboiko/joyodb>>. Acesso em: julho de 2016.
- BOIKO, L. F. S. *Chinês literário: Uma língua escrita*. 2013. Disponível em: <[http://namakajiri.net/letras/tgi/tgi\\_leoboiko.pdf](http://namakajiri.net/letras/tgi/tgi_leoboiko.pdf)>.
- BOLTZ, W. *The origin and early development of the Chinese writing system*. Michigan: American Oriental Society, 1994. (American oriental series). ISBN 9780940490789.
- BOODBERG, P. Some proleptical remarks on the evolution of archaic Chinese. *Harvard Journal of Asiatic Studies*, n. 2, p. 329–372, 1937.
- BOTTÉRO, F. Revisiting the wén 文 and the zì 字: The Great Chinese characters hoax. *Bulletin of the Museum of Far Eastern Antiquities*, Östasiatiska museet, Stockholm, n. 74, p. 14–33, 2004.
- BRANNER, D. P. On early Chinese morphology and its intellectual history. *Journal of the Royal Asiatic Society*, Cambridge University Press, Cambridge, v. 13, n. 1, p. 45–76, 2003. ISSN 13561863, 14740591. Disponível em: <[https://brannerchinese.com/publications/Branner\\_Morphology\\_Intellectual\\_History.pdf](https://brannerchinese.com/publications/Branner_Morphology_Intellectual_History.pdf)>.
- BREEN, J. et al. *The EDICT Dictionary File*. 2015. Disponível em: <<http://www.edrdg.org/jmdict/edict.html>>.
- BREEN, J. et al. *The KANJIDIC/KANJD212 Project*. 2016. Disponível em: <<http://www.edrdg.org/kanjdic/kanjdic.html>>. Acesso em: julho de 2016.
- BRIGHT, W.; DANIELS, P. T. (Ed.). *The world's writing systems*. Oxford: Oxford University Press, 1996.
- BUNKACHŌ. *Heisei 21 Nendo "Kokugo ni kan-suru Seron Chōsa" no Kekka ni tsuite*. Tokyo: Bunkachō [Agência para Assuntos Culturais], 2010. [平成21年度「国語に関する世論調査」の結果について]. Disponível em: <[http://www.bunka.go.jp/kokugo\\_nihongo/yoronchousa/h21/pdf/h21\\_chosa\\_kekka.pdf](http://www.bunka.go.jp/kokugo_nihongo/yoronchousa/h21/pdf/h21_chosa_kekka.pdf)>.
- CAMPOS, H. d.; DANTAS, H. d. L. *Ideograma: lógica, poesia, linguagem*. São Paulo: EDUSP, 1994. ISBN 9788531401701.
- CHAUDHURI, S. K. The *Varṇamālā* in China. *Keiei Kenkyū*, Aichi Gakusen Daigaku, Okazaki, v. 10, n. 3, p. 579–626, 1997. [愛知学泉大学, 『経営研究』]. Disponível em: <<http://ci.nii.ac.jp/lognavi?name=nels&lang=en&type=pdf&id=ART0008579015>>. Acesso em: julho de 2016.
- CHEN, X. et al. Stages in learning to pronounce Chinese characters. *Psychology in the Schools*, Wiley Periodicals, Hoboken, v. 40, n. 1, p. 115–124, 2003.
- CHOMKSY, N. *Aspects of the Theory of Syntax*. Massachusetts: The M.I.T. Press, 1969.

- CHOMSKY, N. *Biolinguistics and the Human Capacity*. 2004. Disponível em: <<http://www.chomsky.info/talks/20040517.htm>>.
- CIA. *CIA World Factbook: Literacy*. 2015. Disponível em: <<https://www.cia.gov/library/publications/the-world-factbook/fields/2103.html>>. Acesso em: 2015.
- CIKOSKI, J. *Notes for a Lexicon of Classical Chinese*. Version 14, draft 7, tweak 144. Georgia: Coprolite, 2011. Disponível em: <<https://www.umass.edu/wsp/resources/books/cikoski/Cikoski-v1.pdf>>. Acesso em: julho de 2016.
- COMRIE, B. (Ed.). *The World's Major Languages*. New York: Oxford University Press, 1987.
- COULMAS, F. *The writing systems of the world*. Hoboken: Wiley, 1989. ISBN 9780631180289.
- DEFRANCIS, J. *The Chinese Language: Fact and Fantasy*. Honolulu: University of Hawai'i Press, 1984. ISBN 9780824810689.
- DEFRANCIS, J. *Visible Speech: The Diverse Oneness of Writing Systems*. Honolulu: University of Hawai'i Press, 1989. ISBN 9780824812072.
- DEFRANCIS, J.; UNGER, J. M. Rejoinder to Geoffrey Sampson, "Chinese script and the diversity of writing systems". *Linguistics*, v. 32, n. 3, p. 549–554, 1994.
- DEHAENE, S. *Reading in the brain: The new science of how we read*. New York: Penguin, 2009.
- DENISOWSKI, P. et al. *CC-CEDICT*. MDGB, 2016. Disponível em: <<https://www.mdbg.net>>. Acesso em: janeiro de 2016.
- DERRIDA, J. *Of grammatology*. 1. ed. Baltimore: Johns Hopkins University Press Baltimore, 1976. Tradução de Gayatri Chakravorty Spivak.
- DONG, Y. *The Prosody and Morphology of Elastic Words in Chinese: Annotations and analyses*. 182 p. Dissertação (Dissertation) — The University of Michigan, Michigan, 2015. Disponível em: <<http://hdl.handle.net/2027.42/116629>>. Acesso em: março de 2016.
- DUANMU, S. Word-length preferences in Chinese: A corpus study. *Journal of East Asian Linguistics*, Springer, v. 21, n. 1, p. 89–114, 2012. Disponível em: <<http://www-personal.umich.edu/~duanmu/CorpusLength2012.pdf>>. Acesso em: março de 2016.
- EDDYSHAW, D. *Too many English loanwords in Japanese?* 2013. Comentário. Disponível em: <<http://languagelog.ldc.upenn.edu/nll/?p=5176#comment-392724>>. Acesso em: julho de 2016.
- EISENSTEIN, S. *Film Form*. New York: Harvest/HBJ, 1949. Tradução de Jay Leyda. ISBN 0-15-630920-3.
- ERBAUGH, M. (Ed.). *Difficult Characters: Interdisciplinary Studies of Chinese and Japanese Writing*. Columbus: Ohio State University, 2002. ISBN 9780874153446.

- FENG, S. Monosyllabicity and disyllabicity in Chinese prosodic morphology. *Macao Journal of Linguistics*, Macao, n. 33, p. 04–19, 2009.
- FENOLLOSA, E. et al. *The Chinese Written Character as a Medium for Poetry: A Critical Edition*. Bronx: Fordham University Press, 1936. ISBN 9780823228683.
- FERGUSON, C. A. Diglossia. In: HYMES, D. (Ed.). *Language in culture and society*. New York: Harper and Row, 1966. p. 429–439.
- FINKENSTAEDT, T.; WOLFF, D. *Ordered profusion; studies in dictionaries and the English lexicon*. Heidelberg: C. Winter, 1973.
- FIORIN, J. L. (Ed.). *Introdução à Lingüística: Princípios de análise*. 4. ed. São Paulo: Contexto, 2008.
- FISCHER, S. R. *History of Writing*. London: Reaktion Books, 2004.
- FRELLESVIG, B. *A History of the Japanese Language*. Cambridge: Cambridge University Press, 2010. ISBN 9780521653206.
- FUYUE, A. *Chōchō ga Ippiki Tattan Kaikyō wo watatte-itta*. 1929. [安西冬衛、てふてふが一匹韃靼海峡を渡つて行つた] (Uma borboleta cruzando o Estreito da Tartária e—foi). Originalmente publicado em: *Gunkan Mari*. Disponível em: <<https://kotobakan.jp/makoto/makoto-5204>>. Acesso em: julho de 2016.
- GB: ZHŌNGHUÁ RÉNMÍN GÒNGHÉGUÓ GUÓJIĀ BIĀOZHŪN [中华人民共和国国家标准]. *Hànyǔ pīnyīn zhèng cífǎ jīběn guīzé* [汉语拼音正词法基本规则]. Beijing, 2012. “Padrão Nacional Da República Popular da China: Regras Básicas da Ortografia Hànyǔ Pīnyīn”. Disponível em: <<http://www.moe.gov.cn/ewebeditor/uploadfile/2012/08/21/20120821100233165.pdf>>. Acesso em: julho de 2016.
- GLEITMAN, L. R.; NEWPORT, E. L. The invention of language by children: Environmental and biological influences on the acquisition of language. *An invitation to cognitive science*, MIT Press, Cambridge, MA, v. 1, p. 1–24, 1995.
- GONÇALVES, C. A.; ANDRADE, K. E.; ALMEIDA, M. L. L. de. Se a macumba é para o bem, então é boacumba: análise morfoprosódica e semântico-cognitiva das substituições sublexicais em português. *Linguística*, Universidade Federal do Rio de Janeiro, Rio de Janeiro, v. 6, n. 2, 2010.
- GOTTLIEB, N. *Kanji Politics: Language policy and Japanese script*. New York: Columbia University Press, 1995. (Japanese Studies). ISBN 0-7103-0512-5.
- GOULD, S. J.; VRBA, E. S. Exaptation—a missing term in the science of form. *Paleobiology*, Paleontological Society, v. 8, n. 1, p. 4–15, 1982. ISSN 00948373, 19385331. Disponível em: <<http://www.jstor.org/stable/2400563>>.
- GROSS, R. *Konjaku no Nihon Meimei ni mirareru Patān*. Akita Daigaku, 2015. [今昔の日本命名に見られるパターン]. Disponível em: <[http://www.pcix.akita-u.ac.jp/inter/pdf\\_class/report\\_2015\\_03.pdf](http://www.pcix.akita-u.ac.jp/inter/pdf_class/report_2015_03.pdf)>. Acesso em: julho de 2016.
- HABEIN, Y. S. *Decoding Kanji*. Tokyo: Kodansha, 2000.

- HALLIDAY, M. A. K. *Halliday's Introduction to Functional Grammar*. 4. ed. New York: Routledge, 2014.
- HANDEL, Z. Logography and the classification of writing systems: a response to unger. *Scripta*, The Hunmin Jeongeum Society, Seul, v. 7, p. 109–150, 2015.
- HANNAS, W. C. *Asia's Orthographic Dilemma*. Honolulu: University of Hawai'i Press, 1997. ISBN 9780824818920.
- HASHIMOTO, M.; SUZUKI, T.; YAMADA, H. *Kanji Minzoku no Ketsudan: Kanji no mirai ni mukete*. Tokyo: Taishūkan Shoin, 1987. [漢字民族の決断]。 (A Decisão dos Povos [usuários de] Kanji).
- HASHIMOTO, Y. *Nihongo no Nazo wo Toku: Saishin gengogaku q&a*. Tokyo: Shinchōsha, 2016. [橋本陽介、『日本語の謎を解く：最新言語学Q&A』] .
- HASPELMATH, M. The growth of affixes in morphological reanalysis. In: \_\_\_\_\_. *Yearbook of Morphology 1994*. Rotterdam: Springer, 1994. p. 1–29. Disponível em: <[http://link.springer.com/chapter/10.1007%2F978-94-017-3714-2\\_1](http://link.springer.com/chapter/10.1007%2F978-94-017-3714-2_1)>. Acesso em: julho de 2016.
- HASPELMATH, M.; SIMS, A. D. *Understanding morphology*. London: Arnold, 2002.
- HEARN, L. *Glimpses of Unfamiliar Japan*. New York: Houghton, Mifflin, 1895. Disponível em: <<https://archive.org/details/glimpsesunfamil09heargooq>>. Acesso em: 2015.
- HEPBURN, J. C. *Japanese-English and English-Japanese Dictionary*. Shanghai: American Presbyterian Mission Press, 1872.
- HERMAN, J. *Vulgar Latin*. Pennsylvania: Penn State Press, 2000.
- IPA. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press, 1999.
- ISO. *ISO/IEC 8859-1:1998—Latin alphabet No. 1*. 1998. Disponível em: <[http://www.iso.org/iso/catalogue\\_detail?csnumber=28245](http://www.iso.org/iso/catalogue_detail?csnumber=28245)>. Acesso em: julho de 2015.
- JJI SHINBŌ KATEI-BU. *Tōkyō Meibutsu Tabē-arū Ki*. Shōwa Dō Shobō, 1929. [時事新報家庭部(編), 『東京名物食べある記』] . Disponível em: <<http://dl.ndl.go.jp/info:ndljp/pid/1176069>>. Acesso em: julho de 2016.
- JOYCE, T. Modeling the Japanese mental lexicon: Morphological, orthographic and phonological considerations. *Advances in Psychology Research*, Nova, New York, v. 31, 2004. Disponível em: <[http://faculty.tama.ac.jp/joyce/papers/Joyce,%20Terry%20\(2004\)%20\[Nova\].pdf](http://faculty.tama.ac.jp/joyce/papers/Joyce,%20Terry%20(2004)%20[Nova].pdf)>. Acesso em: julho de 2016.
- JOYCE, T. The significance of the morphographic principle for the classification of writing systems. *Written Language & Literacy*, John Benjamins Publishing Company, v. 14, n. 1, p. 58–81, 2011.
- JOYCE, T. *Japanese writing system (JWS)*. 2012. Disponível em: <<http://faculty.tama.ac.jp/joyce/jws.html>>.

- KAKIZAKI, M. *Green Blood*. Tokyo: Kōdansha, v. 1, 2011.
- KAMIYA, Y. *No Game No Life*. Tokyo: Media Factory, 2012. 【榎宮 祐, 『ノーゲーム・ノーライフ 1』】. ISBN 978-4-8401-4546-6.
- KARLGRÉN, B. *Analytic Dictionary of Chinese and Sino-Japanese*. New York: Dover, 1923.
- KAYAMOTO, Y.; YAMADA, J.; TAKASHIMA, H. The consistency of multiple-pronunciation effects in reading: The case of Japanese logographs. *Journal of Psycholinguistic Research*, Kluwer, v. 27, n. 6, p. 619–637, 1998. ISSN 0090-6905. Disponível em: <<http://dx.doi.org/10.1023/A%3A1023227904895>>.
- KENNEDY, G. A. The monosyllabic myth. *Journal of the American Oriental Society*, American Oriental Society, Massachusetts, v. 71, n. 3, p. 161–166, 1951. ISSN 00030279. Disponível em: <<http://www.jstor.org/stable/595185>>. Acesso em: fevereiro de 2016.
- KENNEDY, G. A. Fenollosa, Pound and the Chinese character. *Yale Literary Magazine*, Yale University, New Haven, v. 126, n. 5, p. 24–36, 1958.
- KENNEDY, G. A. *Selected works*. New Haven: Far Eastern Publications, Yale University, 1964.
- KESS, J. F.; MIYAMOTO, T. *The Japanese Mental Lexicon: Psycholinguistic Studies of Kana and Kanji Processing*. Philadelphia, PA: John Benjamins, 1999. ISBN 1556197616. Disponível em: <<http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/1556197616>>. Acesso em: julho de 2016.
- KOIKE, M. Genjitsu-sō. In: \_\_\_\_\_. *Nihongo no Tanpen-Shōsetu*. New York: Penguin Group (USA), 2011. ISBN 9780143118336. [小池昌代、『げんじつ荘』。『日本語の短篇小説』に。] (*A Mansão Genjitsu*. In: *Short Stories in Japanese*.)
- KUBOZONO, H. *Handbook of Japanese Phonetics and Phonology*. Berlin: de Gruyter, 2015.
- KUDŌ, T. *MeCab: Yet Another Part-of-Speech and Morphological Analyzer*. 2013. Disponível em: <<http://taku910.github.io/mecab/>>.
- KUHN, T. *The structure of scientific revolutions*. Chicago: University of Chicago Press, 1996. ISBN 9780226458083.
- KYŌGOKU, N. Hakuzōsu. In: \_\_\_\_\_. *Kōsetsu Hyakumonogatari*. Tokyo: Kadokawa, 1999. [京極夏彦、『白蔵主』。『巷説百物語』より。] (*O Monge Hakuzōsu*. In: *Cem Contos Populares*).
- KYŌGOKU, N. *Natsuhiko Kyōgoku ga Shuppangyō-kai no “Kōzō-Kaikaku” wo Unagasu*. Adobe, 2003. [京極夏彦が出版業界の「構造改革」を促す。] (Natsuhiko Kyōgoku clama por uma “restruturação” do mundo publicitário). Disponível em: <<http://www.adobe.com/jp/print/features/kyogoku/>>. Acesso em: julho de 2016.
- LABOV, W. On the mechanism of linguistic change. *Georgetown monographs on language and linguistics*, Georgetown University Press, Washington, v. 18, p. 91–114, 1965.
- LAMARRE, T. *Uncovering Heian Japan: an archaeology of sensation and inscription*. Durham: Duke University Press, 2000. ISBN 9780822325185.

- LEIPZIG, DEPARTAMENTO DE LINGUÍSTICA DA UNIVERSIDADE. *The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses*. 2008. Disponível em: <<http://www.eva.mpg.de/lingua/pdf/LGR08.02.05.pdf>>.
- LENG, Y.; WEI, Y. et al. *Zhōnghuá Zihǎi* [中华字海]. Pequim: Zhōngguó yǒuyì chūbǎn gōngsī [中国友谊出版公司], 1994.
- LINK, P. *An anatomy of Chinese: Rhythm, metaphor, politics*. Cambridge: Harvard University Press, 2013.
- LOOKWISE. *Gogen Yurai Jiten*. Lookwise, 2016. [語源由来辞典]. Disponível em: <<http://gogen-allguide.com/>>. Acesso em: julho de 2016.
- LURIE, D. B. *The Origins of Writing in Early Japan: From the 1st to the 8th Century C.E.* Tese (Doutorado) — Columbia University, 2001. Disponível em: <<http://search.proquest.com/docview/304690137/1378D9395F359639607/2>>.
- LURIE, D. B. Language, writing, and disciplinarity in the critique of the “Ideographic Myth”: Some proleptical remarks. *Language & Communication*, Elsevier, n. 26, p. 250–269, 2006. Disponível em: <<http://www.columbia.edu/~dbl11/Lurie-LangWritingDisc.pdf>>.
- LURIE, D. B. *Realms of Literacy: Early Japan and the History of Writing*. Cambridge: Harvard University Asia Center, 2011.
- LYONS, J. Human language. *Non-verbal communication*, p. 49–85, 1972.
- MAIR, V. *Polysyllabic characters in Chinese writing*. 2011. Disponível em: <<http://linguagelog ldc.upenn.edu/nll/?p=3330>>. Acesso em: julho de 2015.
- MAIR, V. *Polysyllabic characters revisited*. 2015. Disponível em: <<http://linguagelog ldc.upenn.edu/nll/?p=19548>>. Acesso em: julho de 2015.
- MAIR, V. *Character amnesia redux*. 2016. Disponível em: <<http://linguagelog ldc.upenn.edu/nll/?p=25291>>. Acesso em: julho de 2016.
- MAIR, V. *Which is worse?* 2016. Discussão nos comentários. Disponível em: <<http://linguagelog ldc.upenn.edu/nll/?p=23645>>. Acesso em: julho de 2016.
- MAIR, V. H. What is a Chinese “dialect/topolect”? reflections on some key Sino-English linguistic terms. *Sino-Platonic Papers*, University of Pennsylvania, Philadelphia, n. 29, p. 1–31, setembro 1991. Disponível em: <[http://sino-platonic.org/complete/spp029\\_chinese\\_dialect.pdf](http://sino-platonic.org/complete/spp029_chinese_dialect.pdf)>. Acesso em: julho de 2016.
- MAIR, V. H. Script and word in Medieval Vernacular Sinitic. *Journal of the American Oriental Society*, American Oriental Society, Massachusetts, v. 112, n. 2, p. 269–278, 1992. ISSN 00030279. Disponível em: <<http://www.jstor.org/stable/603705>>.
- MAIR, V. H. Buddhism and the rise of the written vernacular in East Asia: the making of national languages. *The Journal of Asian Studies*, Association for Asian Studies, Ann Arbor, v. 53, n. 3, p. 707–751, 1994. ISSN 00219118. Disponível em: <<http://www.jstor.org/stable/2059728>>.

- MARTINET, A. Double articulation as a criterion of linguisticity. *Language Sciences*, v. 6, n. 1, p. 31–38, 1984. ISSN 0388-0001. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0388000184800030>>.
- MATSUNAGA, S. *The Linguistic and Psycholinguistic Nature of Kanji: Do Kanji Represent and Trigger only Meanings?* Tese (Doutorado) — University of Hawai'i, Honolulu, August 1994.
- MEXT. *Tōyō Kanji-hyō*. 1946. [当用漢字]。Pronunciamento do Gabinete do Governo Japonês (内閣国事) .
- MEXT. *Jōyō Kanji-hyō*. 1981. [常用漢字表]。Pronunciamento do Gabinete do Governo Japonês (内閣国事) .
- MEXT. *Jōyō Kanji-hyō*. 2010. [常用漢字表]。Pronunciamento do Gabinete do Governo Japonês (内閣国事) . Disponível em: <[http://www.mext.go.jp/b\\_menu/hakusho/nc/k19811001001/k19811001001.html](http://www.mext.go.jp/b_menu/hakusho/nc/k19811001001/k19811001001.html)>. Acesso em: 2015.
- MEXT. *Revised Proposal of HENTAIGANA*. Japan, 2016.
- MILLER, R. A. *The Japanese Language*. Chicago: The University of Chicago Press, 1967. ISBN 0266527174.
- MIYAJI, Y.; KAI, M. “*Nihongogaku*” Tokushū Tēma Betsu Fairu: Goi 3: Goshuron / Wago. Tokyo: Meiji Shoin, 2008. ISBN 987-4-625-43410-5.
- MIYAKE, M. H. *Old Japanese: a phonetic reconstruction*. New York: Routledge, 2003. ISBN 0415305756.
- MORI, S. *Keitaisokaiseki (Hinshi Suitei)*. 2011. Disponível em: <<http://plata.ar.media.kyoto-u.ac.jp/mori/research/topics/PST/>>.
- MURAKAMI, H. Sidney no green street. In: \_\_\_\_\_. *Hajimete no Bungaku Murakami Haruki*. Tokyo: Bungen Shunshū, 2006.
- MYERS, J. *Knowing Chinese character grammar*. 2015. Disponível em: <<http://ling.auf.net/lingbuzz/002562>>. Acesso em: julho de 2015.
- NAKAMURA, K. et al. Subliminal convergence of kanji and kana words: Further evidence for functional parcellation of the posterior temporal cortex in visual word perception. *Journal of Cognitive Neuroscience*, Massachusetts Institute of Technology, v. 17, n. 6, p. 954–968, 2005.
- NICHOLS, C. *How to Read Egyptian Hieroglyphs: For High School Students in Grades 9 Through 12*. [S.l.]: CreateSpace Independent Publishing Platform, 2008. ISBN 9781438264684.
- NICO NICO PEDIA. *Niji*. 2015. [虹裏] . Disponível em: <<http://dic.nicovideo.jp/a/%E8%99%B9%E8%A3%8F>>. Acesso em: julho de 2016.
- NICO NICO PEDIA. *Niji*. 2016. [超] . Disponível em: <<http://dic.nicovideo.jp/a/%E8%B6%85>>. Acesso em: julho de 2016.

- NINJAL. *Gakujutsu Jōhō Kōkan-yō Hentaigana*. 2016. [国立国語研究所、『学術情報交換用変体仮名』] (Instituto Nacional de Língua Japonesa e Linguística, “Hentaigana para fins de comunicação na informática”). Disponível em: <<http://kana.ninjal.ac.jp/>>. Acesso em: julho de 2016.
- NOSAKA, A. *Hotaru no Haka*. Tokyo: Sanyusha, 1967. [野坂 昭如、『火垂るの墓』] .
- NPC: QUÁNGUÓ RÉNMÍN DÀIBIǎO DÀHUÌ [全國人民代表大會] . *Quánrìzhì Yǐwù Jiàoyù Yǔwén Kèchéng Biāozhǔn* [全日製義務教育語文課程標準] . 2001. “Congresso Popular Nacional da China: Norma do currículo linguístico compulsório integral”. Disponível em: <<http://goo.gl/DuI7rT>>. Acesso em: julho de 2015.
- NUNBERG, G. *The Linguistics of Punctuation*. Stanford: Center for the Study of Language and Information, 1990. (CSLI Lecture Notes). ISBN 9780937073469.
- OGIHARA, A.; KŌNOSU, H. (Ed.). *Nihon Koten Bungaku Zenshū: Kojiki; jōdai Kayō*. Tokyo: Shogakukan, 1973. [荻原浅男; 鴻巣隼雄。『日本古典文学全集 1 : 古事記 上代歌謡』] .
- OKAMOTO, K. *Koten Nihongo no Sekai: Kanji ga tsukuru Nihon*. Tokyo: Kenkyūsha, 2007. [『古典日本語の世界 : 漢字がつくる日本』] .
- ONMYŌZA. *Onmyō-za Kōshiki Homepage: Shin release jōhō*. 2015. [陰陽座公式庵頁 : 最新リリース情報] . Disponível em: <<https://www.onmyo-za.net/discography/index.html>>. Acesso em: julho de 2016.
- OPENCLIPART.ORG. *Openclipart*. 2016. Disponível em: <<https://openclipart.org>>. Acesso em: janeiro de 2016.
- O’NEILL, T. Xu Shen’s scholarly agenda: A new interpretation of the postface of the Shuowen Jiezi. *Journal of American Oriental Society*, American Oriental Society, Massachusetts, v. 133, n. 3, p. 413–440, 2013. ISSN 00030279. Disponível em: <<http://www.jstor.org/stable/10.7817/jameroriesoci.133.3.0413>>. Acesso em: janeiro de 2016.
- PACKARD, J. L. *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge: Cambridge University Press, 2000.
- PERNISS, P.; THOMPSON, R.; VIGLIOCCO, G. Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology*, Frontiers, Bruxelles, v. 1, n. 227, 2010. ISSN 1664-1078. Disponível em: <[http://www.frontiersin.org/language\\_sciences/10.3389/fpsyg.2010.00227/abstract](http://www.frontiersin.org/language_sciences/10.3389/fpsyg.2010.00227/abstract)>.
- PINKER, S. *The language instinct: How the mind creates language*. New York: HarperCollins, 2010.
- PONCEAU, P. S. D. *A dissertation on the nature and character of the Chinese system of writing in a letter to John Vaughan*. Philadelphia: Esq. American Philosophical Society, 1838. Disponível em: <<http://archive.org/details/dissertationnatu00duporich>>.
- PRINCETON UNIVERSITY. *About WordNet*. Princeton University, 2010. Acessado via interface Python NLTK. Disponível em: <<http://wordnet.princeton.edu>>. Acesso em: Julho de 2016.

- PULLUM, G. *The Awful Chinese Writing System*. 2016. Disponível em: <<http://chronicle.com/blogs/linguafranca/2016/01/20/the-awful-chinese-writing-system>>. Acesso em: julho de 2016.
- PYTHON SOFTWARE FOUNDATION. *Python*. 2015. Disponível em: <<https://www.python.org/>>.
- R FOUNDATION FOR STATISTICAL COMPUTING. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2015. Disponível em: <<https://www.R-project.org>>.
- R25. “(笑), “藁”, “w” *Hibi Shinka-shita “Netto Surangu”-Hen*. 2010. [「(笑)」 「藁」 「w」 日々進化した「ネットスラング」編] (“Compilação da evolução diária da gíria de Internet: (笑), 藁, ‘w’”). Disponível em: <<http://r25.jp/topic/00001014/>>. Acesso em: julho de 2016.
- ROBERT, J.-N. Hieroglossia: A proposal. *Bulletin of the Nanzan Institute for Religion & Culture*, Nanzan Institute for Religion and Culture, Nagoya, n. 30, p. 25–48, 2006.
- ROBINSON, A. *The story of writing: Alphabets, hieroglyphs & pictograms*. London: Thames & Hudson, 2007.
- ROGERS, H. *Writing Systems: A Linguistic Approach*. Hoboken: John Wiley & Sons, 2004. (Blackwell Textbooks in Linguistics). ISBN 9780631234647.
- SAITŌ, T. *Gakkō de wa oshiete kurenai Nihongo no Jugyō*. Kyoto: PHP, 2014. [齋藤 孝、『学校では教えてくれない日本語の授業』]. ISBN 9784569819211.
- SAMPSON, G. Chinese script and the diversity of writing systems. *Linguistics*, n. 32, p. 117–132, 1994. Disponível em: <<http://www.grsampson.net/ACsa.html>>.
- SANSOM, G. B. *An historical grammar of Japanese*. Oxford: The Clarendon Press, 1928. Disponível em: <<http://archive.org/details/historicalgramma00sansuoft>>.
- SAUSSURE, F. D. *Curso de Linguística geral*. São Paulo: Editora Cultrix, 2008.
- SAUSSURE, F. de et al. *Course in General Linguistics*. New York: Columbia University Press, 2011. ISBN 9780231527958.
- SCHOLTEN, D.; GÍSLADÓTTIR, R. *Woher kommt das Wort Religion?* München: Belles Lettres, 2015. Disponível em: <<http://www.belleslettres.eu/artikel/religion-religare-relegere.php>>. Acesso em: julho de 2016.
- SCHUESSLER, A. *ABC etymological dictionary of Old Chinese*. Honolulu: University of Hawai’i Press, 2007.
- SCHUESSLER, A. *Minimal Old Chinese and later Han Chinese: a companion to Grammata Serica Recensa*. Honolulu: University of Hawaii Press, 2009.
- SEARS, R. H. *Chinese Etymology*. 2013. Disponível em: <<http://chineseetymology.org>>. Acesso em: janeiro de 2016.
- SEELEY, C. Special issue: Aspects of the Japanese writing system. *Visible Language*, University of Cincinnati, Cincinnati, v. 18, n. 3, p. 219–228, 1984. Editorial introdutório. Disponível em: <<http://visiblelanguagejournal.com/issues/issue/71/>>.

- SEELEY, C. *A history of writing in Japan*. Leiden: E.J. Brill, 1991. (Brill's Japanese studies library). ISBN 9789004090811.
- SHIBATANI, M. *The Languages of Japan*. Cambridge: Cambridge University Press, 1990. ISBN 0521360706.
- SHIBLES, W. A. IPA transliteration of Chinese romanization systems. *Sino-Platonic Papers*, University of Pennsylvania, Philadelphia, n. 52, p. 1–20, novembro 1994. Disponível em: <[http://sino-platonic.org/complete/spp052\\_chinese\\_ipa.pdf](http://sino-platonic.org/complete/spp052_chinese_ipa.pdf)>. Acesso em: julho de 2016.
- SHŪEISHA. *Shūkan Shonen Jump 2012 No. 31*. Tokyo: Shūeisha, n. 31, julho 2012. [週刊少年ジャンプ No.31]。ASIN B008GOMRN0.
- SIBI/USP, DEPARTAMENTO TÉCNICO. *Diretrizes para apresentação de dissertações e teses da USP: documento eletrônico e impresso: Parte I (ABNT)*. São Paulo: Sistema Integrado de Bibliotecas da USP, 2009. Disponível em: <[http://www.teses.usp.br/index.php?option=com\\_content&view=article&id=52&Itemid=67](http://www.teses.usp.br/index.php?option=com_content&view=article&id=52&Itemid=67)>.
- SMITH, J. S. Japanese writing. In: BRIGHT, W.; DANIELS, P. T. (Ed.). *The world's writing systems*. Oxford: Oxford University Press, 1996.
- SNOW, D. Diglossia in East Asia. *Journal of Asian Pacific Communication*, John Benjamins, Amsterdam, v. 20, n. 1, p. 124–151, 2010.
- SOOTHILL, W. E. *The Student's Four Thousand 字 and General Pocket Dictionary*. Shanghai: American Presbyterian Mission Press, 1908. Disponível em: <<https://archive.org/details/studentsfourtho00sootgoog>>. Acesso em: agosto de 2015.
- SPROAT, R. W. *A computational theory of writing systems*. Cambridge, Massachusetts: MIT Press, 2000.
- TABATA, A. “Kanji Būmu” no hyōgen to jittai. *Kobe Daigaku Daigakuin Ningen Hattatsu Kankyōgaku Kenkyūka Kenkyū-Kiyō*, Kobe University, Kobe, v. 8, n. 2, p. 41–48, 2015. [田畑 暁生、『「漢字ブーム」の表象と実態』。神戸大学大学院人間発達環境学研究科研究紀要。]。Disponível em: <<http://www.lib.kobe-u.ac.jp/repository/81008823.pdf>>. Acesso em: 2015.
- TAKAHATA, I. *Hotaru no Haka*. Japão: Studio Ghibli, 1988. Animação.
- TAKEUCHI, N. *Bishōjo Senshi Sailor Moon S: Kaguya-hime no Koibito*. Tokyo: Kōdansha, 1994.
- TAKUMI, S. (Direção). *Ghost Trick*. Japão: Capcom, 2010. Software para videogame Nintendo DS. [『ゴースト トリック』]。
- TANG, T. *Hanzi Smatter*. 2015. Disponível em: <<http://hanzismatter.blogspot.com>>. Acesso em: 2015.
- TANIZAKI, J. *Sasameyuki*. Tokyo: Shinchōsha, 1995. [谷崎潤一郎、『細雪』。]。
- TANIZAKI, J. *As Irmãs Makioka*. São Paulo: Estação Liberdade, 2005. Tradução de Leiko Gotoda, Kanami Hirai, Neide Hissae Nagae e Eliza Atsuko Tashiro.

- THE JAPAN TIMES. S. Korea student suicides hit 139 in '12. *The Japan Times*, Tokyo, agosto 2013. Disponível em: <<http://www.japantimes.co.jp/news/2013/08/20/asia-pacific/social-issues-asia-pacific/s-korea-student-suicides-hit-139-in-12/>>. Acesso em: julho de 2016.
- THE LANGUAGE ARCHIVE. *Documentation of Endangered Languages: Taa*. 2015. Disponível em: <<http://dobes.mpi.nl/projects/taa/language/>>. Acesso em: julho de 2015.
- THE UNICODE CONSORTIUM. *The Unicode Standard*. 9.0.0. ed. Mountain View, 2016. Disponível em: <<http://www.unicode.org/versions/Unicode9.0.0/>>. Acesso em: julho de 2016.
- TŌDŌ, A. *Kanji Gogen Jiten*. Tokyo: Gakuto, 1965. [藤堂 明保、『漢字語源辞典』].
- TREYVAULD, M. *Western tofu*. 2015. Disponível em: <[http://no-sword.jp/blog/2015/07/western\\_tofu.html](http://no-sword.jp/blog/2015/07/western_tofu.html)>. Acesso em: julho de 2016.
- TSUJIMURA, N. *An Introduction to Japanese Linguistics*. 2. ed. Oxford: Blackwell Oxford, 2006. ISBN 1405110651.
- TZENG, O. J. et al. Learning to be a conspirator: A tale of becoming a good Chinese reader. In: \_\_\_\_\_. Mahwah: Erlbaum, 1995. p. 227–246.
- UNGER, J. M. *The fifth generation fallacy: Why Japan is betting its future on artificial intelligence*. New York: Oxford University Press, 1987.
- UNGER, J. M. *Literacy and script reform in occupation Japan: Reading between the lines*. New York: Oxford University Press, 1996. ISBN 9780195356380.
- UNGER, J. M. *Ideogram: Chinese Characters and the Myth of Disembodied Meaning*. Honolulu: University of Hawai'i Press, 2004. ISBN 9780824827601.
- VACHEK, J.; LUELSDORFF, P. *Written language revisited*. Philadelphia: J. Benjamins, 1989. ISBN 9789027220646.
- WAKISAKA, G. *Manyôshu: Vereda do poema clássico japonês*. São Paulo: Hucitec, 1992.
- WARREN, J. In *Egyptian the water hieroglyph is pronounced by Egyptologists as /n/ but...?* Yahoo Answers, 2005. Disponível em: <<https://answers.yahoo.com/question/index?qid=20060716160554AAPvqfc>>. Acesso em: fevereiro de 2016.
- WATSUKI, N. *Rurōni Kenshin: Kan no Roku*. Tokyo: Shūeisha, v. 6, 1995. ISBN 978-4088715063.
- WETZEL, L. Types and tokens. In: ZALTA, E. N. (Ed.). *The Stanford Encyclopedia of Philosophy*. Spring 2014. Stanford: [s.n.], 2014.
- WIEGER, L. *Chinese Characters*. Ho-kien fu: Catholic Mission Press, 1927. Tradução de L. Davrout.
- WIKIMEDIA FOUNDATION. *Wikimedia Downloads*. 2015. Disponível em: <<https://dumps.wikimedia.org/>>. Acesso em: agosto de 2015.

WRIGHT, R. *Late Latin and Early Romance in Spain and Carolingian France*. Liverpool: Francis Cairns, 1982.

WYLLYS, R. E. Empirical and theoretical bases of Zipf's law. *Library Trends*, University of Illinois Library School, Champaign-Urbana, v. 30, n. 1, p. 53–64, 1981.

YAMAGUCHI, Y. *Wikipedia no Dēta kara Kōpasu Sakusei*. 2011. 『Wikipediaのデータからコーパス作成』 . Disponível em: <<http://www.ar.media.kyoto-u.ac.jp/members/yohei/wikipedia.html>>.

YAMAI, K. (Direção). *Shin Megami Tensei IV*. Japão: Atlus, 2013. Software para videogame Nintendo DS. [『真・女神転生IV』] .

YODA, T. Literary history against the national frame, or gender and the emergence of Heian kana writing. *Positions: East Asia cultures critique*, Duke University Press, Durham, v. 8, n. 2, p. 465, 2000.

YUSHU, Z.; TINGJING, C. et al. *Kāngxī Zìdiǎn* [康熙字典] . 1716. Disponível em: <<http://www.kangxizidian.com/>>. Acesso em: fevereiro de 2016.

ZESHAN, U. Hand, head, and face: Negative constructions in sign languages. *Linguistic Typology*, De Gruyter Mouton, Konstanz, v. 8, n. 1, p. 1–58, 2004. Disponível em: <<https://www.degruyter.com/abstract/j/lity.2004.8.issue-1/lity.2004.003/lity.2004.003.xml>>. Acesso em: julho de 2016.

ZHOU, X.; MARSLEN-WILSON, W. The nature of sublexical processing in reading Chinese characters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, American Psychological Association, Washington, v. 25, n. 4, p. 819–837, 1999.

ŌSHIRO, T. Kamekō-baka. In: \_\_\_\_\_. *Ōshiro Tatsuhiko Zenshū*. Tokyo: Bensei, 2002. [大城立裕、『亀甲墓』] .



# Apêndices



## APÊNDICE A – Glossário de termos técnicos

- Acento: 1) A propriedade, na maioria das línguas, de dar mais proeminência a uma das sílabas de cada palavra. Essa proeminência pode ser realizada através da articulação fonética, da duração, da intensidade ou da altura; o último denomina-se acento tonal (q.v.). 2) Acento gráfico; o mesmo que diacrítico (q.v.).
- Acento gráfico: O mesmo que diacrítico (q.v.).
- Acento tonal: Propriedade de algumas línguas nas quais a sílaba acentuada (ver acento, 1) é determinada sobretudo pela variação de altura (grave/agudo). Um exemplo é o japonês, onde o acento é marcado por uma queda na altura.
- Alfabeto: Um sistema de escrita baseado no princípio alfabético, viz., em que cada grafema (letra) representa um fonema.
- Alfabeto latino: O conjunto de letras usado para representar o latim; por extensão, os vários sistemas de escrita baseados nesse conjunto, como o português, o espanhol, o inglês etc.
- Alfabeto romano: ver alfabeto latino.
- Altura: Em acústica e auditória, a frequência do som, que leva-o a ser percebido como mais grave ou agudo. Ao contrário do uso da palavra na linguagem cotidiana, “altura” no sentido técnico não se refere à intensidade do som (o volume).
- Afixo: Um morfema preso que se liga a uma base; pode ser um prefixo, sufixo etc.
- Alofonia, alofone: Uma pronúncia alternativa de um fonema (q.v.), que não chega a mudar o sentido.
- Alografia, alógrafo: Uma representação visual alternativa de um grafema (q.v.), que não chega a mudar o sentido; um dos grafos (q.v.) de um grafema.
- Alomorfia, alomorfe: Uma pronúncia alternativa de um morfema (q.v.), muitas vezes considerada como variação condicionada (q.v.).
- Alvéolo: A região do céu da boca atrás dos dentes frontais, semelhante a uma “lombada”. Usada na produção de sons como o /t/, /l/ e /s/.
- Aparelho articulador: O conjunto de órgãos que produz e modula os sons da linguagem: língua, lábios, dentes, cordas vocais, glote etc.

- Articulação, fonológica: Os gestos que o aparelho articulador (cf.) faz para produzir os sons da linguagem.
- Articulação da linguagem: O fato que os enunciados linguísticos são compostos por unidades menores recorrentes. Cp. “Dupla articulação da linguagem”.
- Associação semântica (tipo de caractere):
- Aspiração: Em fonologia, a pronúncia de algumas consoantes acompanhada por um sopro de ar, semelhante ao som do /h/. É um fenômeno fonético no inglês, e fonêmico no mandarim moderno e no grego antigo, dentre outras línguas.
- Base: A parte de uma palavra que não varia quando é flexionada.
- Cabeçalho de seção: Na tradição chinesa, componentes (q.v.) que foram selecionados para organizar dicionários e facilitar a busca de caracteres. O conjunto mais usado são os 214 cabeçalhos do dicionário Kangxi.
- Caractere: O mesmo que grafema (q.v.).
- Caractere chinês: Um grafema morfográfico, originário do sistema de escrita chinês. Pode ser simples, ou um agregado de componentes. Tradução do chinês 漢字 *hànzì* (japonês *kanji*).
- Componente: No contexto dos caracteres chineses, elementos visuais recorrentes que são usados para compô-los.
- Componente fonético: Um componente (q.v.) que serve como pista da pronúncia de um caractere chinês.
- Componente fono-semântico dual: Um componente de um caractere chinês que serve simultaneamente como pista de pronúncia e de sentido.
- Componente semântico: Um componente de um caractere chinês que serve como pista de sentido. Cp. heterografia, determinativo.
- Condicionada, variação: ver “Variação condicionada”.
- Conjugação: flexão (q.v.) de um verbo.
- Contagem de ocorrências: Ver ocorrências.
- Contagem de tipos: Ver tipos.
- Declinação: flexão (q.v.) de um substantivo.

- 
- Determinativo: Marca gráfica que, acrescida a um grafema-base, seleciona uma de suas várias pronúncias (determinativo fonético) ou um de seus vários sentidos (determinativo semântico).
  - Diacrítico: Marca adicional acrescida a um grafo (q.v.) para alterar seu significado, como os acentos gráficos do alfabeto português. Pode ser considerado uma marca gráfica “presa”.
  - Diacronia, diacrônico: Em “tempos diferentes”; de um ponto de vista que leva em conta o desenvolvimento histórico.
  - Diglossia: Situação social na qual uma língua é amplamente usada na escrita mas não na fala.
  - Dupla articulação da linguagem: O fato que os enunciados linguísticos podem ser subdivididos (articulados) em elementos unitários de som e sentido, chamados morfemas; e estes, por sua vez, podem ser subdivididos em unidades menores (fonemas), desprovidas de sentido.
  - Empréstimo homofônico, caractere de:
  - Entonação: Variação na velocidade, altura (nota) e intensidade (volume) na fala.
  - Enunciado: Qualquer emissão linguística provida de sentido; uma frase, oração, período ou fragmento, pronunciado ou escrito em algum contexto faz sentido.
  - Flexão: Uma variação no som e sentido de uma palavra. Por exemplo, “gatas” é a flexão feminina e plural de “gato”.
  - Fone: Unidades de som usadas pela linguagem para compor significantes. São agrupados em fonemas; dois fones que pertençam ao mesmo fonema podem ser trocados um pelo outro, sem alterar o sentido.
  - Fonema: Unidades de som usadas pela linguagem para compor significantes (morfemas, palavras etc.). Ao mudar um fonema, muda-se o significante, e consequentemente o significado. Cada fonema pode ser pronunciado (“realizado”) de maneiras diferentes; cada pronúncia possível é um fone (q.v.).
  - Fonografia: A escrita baseada na representação dos fonemas (q.v.).
  - Fonograma: Um grafema que representa fonemas.
  - Glifo: O mesmo que grafo (q.v.). Termo usado na tecnologia da informática.
  - Grafema: Um significante elementar de um sistema de escrita, capaz de distinguir significado.

- Grafo: Uma marca gráfica de um sistema de escrita. Um grafema (q.v.) pode ter mais de um grafo; e.g. ⟨f⟩ e ⟨f⟩ são grafos do mesmo grafema “F”. Análogo escrito de “fone” (q.v.).
- Heterografia: Quando a escrita diferencia significantes que na fala são idênticos (homófonos).
- Homofonia: Coincidência de sons; pronúncia igual. Diz-se especialmente de dois morfemas de significante idêntico e significado distinto.
- Ideograma: Um grafo que representa diretamente uma ideia, sem fazer referência a palavras ou morfemas específicos.
- Lenição: Ver “abrandamento”.
- Letra: Um grafema de um alfabeto (q.v.).
- Livre: Dito de um elemento que pode ser empregado sozinho; aplica-se a morfemas, grafemas, componentes etc. Antônimo de preso.
- Livre, variação: ver “Variação livre”.
- Logografia: A escrita baseada na representação das palavras. Não existem sistemas puramente logográficos conhecidos. A expressão por vezes é usada como sinônimo de morfografia (q.v.).
- Mora: Unidade abstrata de duração ou quantidade de uma sílaba. Uma sílaba curta (como o japonês o “rabo”) possui uma mora, enquanto que uma sílaba longa (japonês ō “rei”) possui duas.
- Morfema: Uma unidade mínima de sentido; uma sequência de sons que possui um sentido, tal que nenhuma subsequência possua sentido. E.g. a palavra “anticorpo” pode ser dividida em dois morfemas, “anti-” e “corpo”; não é possível dividir “anti-” em componentes providos de sentido.
- Morfografia: Um sistema de escrita baseado no princípio morfográfico, que consiste em representar morfemas por meio de grafemas.
- Morfograma: Um grafema que representa um morfema.
- Morfossilabograma: Um grafema que representa ao mesmo tempo uma sílaba e um morfema; uso típico dos caracteres chineses.
- Morfofonêmico: O nível intermediário entre o fonêmico e o morfêmico; um processo que envolva a interface entre os dois níveis, tipicamente a seleção de pronúncias alternativas de um morfema (alomorfes).

- Ocorrências, contagem de: Uma contagem que leva em conta as ocorrências de itens repetidos. Por exemplo, a frase “guerreiros com guerreiros” possui 3 palavras na contagem de ocorrências, sendo que “guerreiros” ocorre duas vezes. Antônimo de contagem de tipos.
- Opaco: Diz-se de uma composição cujo sentido não pode ser previsto a partir de seus morfemas constituintes. Por exemplo, “amor-perfeito” não é um tipo de amor, mas uma espécie de flor. Antônimo de transparente.
- Ortografia: A grafia prescrita como correta em determinado sistema de escrita; a representação gráfica de uma palavra em grafemas, tal como preestabelecida por uma comunidade letrada.
- Palatalização: Mudança fonética que aproxima a articulação de um fonema ao palato, a região central do céu da boca. No português brasileiro, inclui a alofonia de /ti,di/ como [tʃi, dʒi]. No japonês, inclui a alofonia de /si,zi,ti/ como /ɕi, ʑi, tɕi/. Historicamente, nas línguas latinas, inclui a mudança do /k/ para /s/ antes de /i,e,ɛ/—o que, por inércia da ortografia, teve como consequência a leitura da letra ⟨c⟩ como /s/ nesse contexto.
- Paradigma, científico: Um trabalho científico que é adotado como modelo por um grupo social, que passa a investigar as questões propostas por aquele modelo seguindo seus métodos.
- Paradigma, morfológico: Uma série de flexões compartilhadas por uma classe de palavras. Por exemplo, os verbos portugueses regulares terminados em “-ar” partilham um paradigma: “cantar”, “canto”, “cantemos”, etc.; “beijar”, “beijo”, “beijemos” etc.
- Pictograma: Um morfograma (q.v.) baseado na representação pictórica do sentido; um desenho do morfema.
- Pluraridade: Em sistemas de escrita, uma sequência de elementos.
- Plurifonia: Quando um grafema representa uma sequência de fonemas, e essa sequência é maior do que o princípio do sistema supõe. Por exemplo, no português normalmente as letras representam um só fonema; mas a letra ⟨x⟩, quando denota /ks/, representa dois.
- Plurigrafia: Quando uma sequência de grafemas é usada para representar algo que, pelo princípio do sistema, precisaria só de um. Por exemplo, no português as duas letras ⟨nh⟩ representam um só som, /ɲ/.

- Plurimorfemia: Quando um morfograma (q.v.) representa mais de um morfema. Por exemplo, em japonês o morfograma 雷 pode representar a palavra *kami-nari* “trovão”, que é composta por dois morfemas, “voz” (*nari*) dos “deuses” (*kami*) (a grafia morfológicamente regular seria 神鳴).
- Preso: Diz-se de um elemento que só pode aparecer em conjunto com outros. Aplica-se a morfemas, grafemas, componentes etc. Antônimo de livre.
- Polissemia: Tipo de polivalência (q.v.) em que há mais de um sentido possível para o mesmo significante.
- Polivalência: Em sistemas de escrita, quando existe mais de uma possibilidade de representação ou leitura.
- Posterior, vogal: ver “vogal posterior”.
- Prefixo: Um morfema preso que se acopla antes de uma base.
- Preso, morfema ou posição: ver “morfema preso”
- Princípio alfabético: Método de representação escrita no qual cada grafema corresponde exatamente a um fonema; uma variante do princípio fonográfico (q.v.) onde a correspondência é de um-para-um.
- Princípio fonográfico: Método de representação escrita no qual os grafemas representam fonemas.
- Princípio morfográfico: Método de representação escrita no qual os grafemas representam morfemas.
- Profundidade ortográfica: A propriedade de algumas escritas fonográficas (q.v.) onde a representação não se atém somente aos fonemas, mas também tem influência da morfologia. Em uma escrita fonográfica profunda, alomorfes (q.v.) de pronúncia distinta podem ser grafados da mesma maneira.
- Quantidade (silábica): Propriedade de algumas línguas, como o japonês, latim e italiano, que distinguem fonemas por sua duração. Por exemplo, em japonês *bushi* é “guerreiro”, *bussshi*, “material”; *obasan* é “tia”, *obaasan*, “avó”, etc.
- Radical: 1. O mesmo que base (q.v.). Distinguir de raiz. 2. Uma má tradução para os conceitos de cabeçalho de seção (q.v.) e componente semântico (q.v.), baseada em uma analogia inválida com o sentido 1.
- Raiz: Uma base (q.v.) composta por apenas um morfema.

- Rebus: Tipo de signo em que um desenho é convertido em uma palavra, e depois tomado como fonograma (q.v.) para o som daquela palavra. Por exemplo,  representando a palavra “soldado”.
- Romanização: Transcrição de um sistema de escrita qualquer, como o chinês ou o japonês, em letras do alfabeto romano (q.v.).
- Semantograma: Um grafo que denota diretamente um sentido ou conceito, sem necessariamente representar o significante
- Semiótica: O estudo dos signos.
- Significado: Aquilo que um signo denota, ou “toma o lugar de”.
- Significante: A forma externa de um signo, que distingue-os dos demais.
- Signo: Um par de significante e significado (q.v.).
- Sinal diacrítico: ver “diacrítico”.
- Silabário: Um conjunto de grafos que representam sílabas.
- Sincronia, sincrônico: “No mesmo tempo”; do ponto de vista do momento presente. Esp. dito da linguagem tal como ela é adquirida pelo falante, sem fazer referência à etimologia ou história da língua.
- Sinítico: O mesmo que “chinês”; no contexto da língua japonesa, diz-se de morfemas de origem chinesa (*on*) ou palavras por eles formados (*kango*).
- Sinograma: O mesmo que caractere chinês.
- Sufixo: Um morfema preso que se acopla depois de uma base.
- Tipo, contagem de: A contagem que não leva em conta o número de ocorrências dos itens. Por exemplo, a frase “guerreiros com guerreiros” possui 2 palavras na contagem de tipos: “guerreiros” e “com”. Antônimo de contagem de ocorrências.
- Tom, lexical: Características de línguas como a chinesa na qual a variação de altura (grave/agudo) faz distinção fonêmica, isto é, diferencia palavras. Distinguir de acento tonal.
- Transparente: Diz-se de uma composição cujo sentido pode ser deduzido a partir dos morfemas que a constituem. Por exemplo, “para-lama” é um objeto para parar a lama, e “maçã-verde” é um tipo de maçã que é verde. Antônimo de opaco.

- **Varição condicionada:** Uma pronúncia alternativa que é totalmente previsível a partir de seu contexto fonológico. Por exemplo, em português, o pronome “o” (“beije-o”) é pronunciado como “lo” sempre que se segue a um verbo que termina em “r” (“beijá-lo”).
- **Varição livre:** Uma variação que não é condicionada (q.v.). Muitas vezes a variação dita “livre” pode ser prevista a partir de fatores externos, como registro ou dialeto, ou ainda por análise estatística, etc.; mas ela não tem regras simples como a variação condicionada.
- **Vernáculo:** Em uma situação de diglossia (q.v.), a língua falada local, em oposição à língua escrita de prestígio.
- **Vogal posterior:** Vogal pronunciada com elevação da língua na parte de trás da boca: no português, /u, o, ɔ/.

## APÊNDICE B – Glossário de termos estrangeiros

*cí*: Palavra chinesa (em oposição a morfema).

*default*: “Por omissão”; dentre várias alternativas, aquela que é escolhida quando nenhuma seleção específica foi feita.

*furigana*: Glosas em *kana*, fonográficas, adicionadas aos morfogramas *kanji*. Literalmente “*kana* salpicado”.

*hànzì*: Caractere chinês, esp. tal como usado nas línguas chinesas.

*hebon-shiki*: Padrão Hepburn de romanização japonesa.

*hiragana*: Um dos conjuntos de caracteres *kana*, baseado em cursivização de *kanji*.

*kana*: 1. (Uso moderno) Fonogramas moraicos do japonês. 2. (Uso antigo) Um *kanji* usado como fonograma.

*kanbun*: Língua chinesa literária (em oposição à falada). Lit. “texto chinês”.

*kanji*: Caractere chinês, esp. tal como usado no japonês.

*katakana*: Um dos conjuntos de caracteres *kana*, baseado na abreviação de *kanji*.

*kun*: “Explicação”; morfemas japoneses nativos atribuídos como leituras de *kanji*, por meio do processo de tradução.

*on*: “Som”; morfemas sinóticos inseridos no japonês por meio escrito, na forma de leituras de *kanji* “pelo som”.

*zì*: 1. Caractere. 2. Caractere chinês, *hànzì*. 3. Por extensão, morfema (que são grafados como *zì*).



## APÊNDICE C – A transcrição do mandarim padrão *pīnyīn*

Na tradição fonológica chinesa, os fonemas são classificados por sua posição: inicial ou final (que, em termos linguísticos, correspondem ao “ataque” e à “rima” da sílaba). Assim separamo-los abaixo.

No caso das vogais, é possível analisar [e], [o], [ɤ] e [ə] como variações condicionadas do mesmo fonema /ə/; e o mesmo vale para [ɛ/æ] e [a]. Optamos por transcrever a variação fonética, em um nível confortável para o falante do português. Omitimos as finais que são simples composições (como *-ua* = *-u* + *-a*, etc.).

(GB, 2012; SHIBLES, 1994)

**Tabela 56 – Notação *pīnyīn* e sua pronúncia**

Pīnyīn	IPA	Nota
<b>Iniciais</b>		
<i>b</i>	/p/	Como em “pai”, não como em “bem”.
<i>p</i>	/p <sup>h</sup> /	“P” aspirado, acompanhado de sopro de ar.
<i>m</i>	/m/	
<i>f</i>	/f/	
<i>d</i>	/t/	Como em “tom”, não como em “dom”.
<i>t</i>	/t <sup>h</sup> /	“T” aspirado, acompanhado de sopro de ar.
<i>n</i>	/n/	
<i>l</i>	/l/	
<i>g</i>	/k/	Como em “casa”, não como em “gato”.
<i>k</i>	/k <sup>h</sup> /	“K” aspirado, acompanhado de sopro de ar.
<i>h</i>	/x/	Semelhante ao “R” carioca, ou ao escocês <i>loch</i> . Fricativa velar não-vozeada.
<i>j</i>	/tɕ/	Semelhante ao português “tcham”, mas com a ponta da língua atrás dos dentes de baixo; /t/ mais o som do japonês <i>shin</i> , /ɕ/. Africada alvéolo-palatal não-vozeada.
<i>q</i>	/tɕ <sup>h</sup> /	Como o <i>j</i> acima, mas aspirado (seguido de sopro de ar). Africada alvéolo-palatal não-vozeada aspirada.
<i>x</i>	/ɕ/	Como no japonês <i>shin</i> . Semelhante ao português “xícara”, mas com a ponta da língua atrás dos dentes de baixo. Fricativa alvéolo-palatal.

<i>zh</i>	/tʃ/	Semelhante ao português “tcham”, mas com o dorso da língua mais para trás. Africada pós-alveolar não-vozeada.
<i>ch</i>	/tʃʰ/	Semelhante ao <i>z</i> acima, mas aspirado (acompanhado de sopro de ar). Africada pós-alveolar não-vozeada aspirada.
<i>sh</i>	/ʃ/	Semelhante ao português “xícara”, mas com o dorso da língua mais para trás. Fricativa pós-alveolar não-vozeada.
<i>r</i>	/ʐ/ ou [ɹ]	Semelhante ao português “já”, mas com dorso da língua mais para trás; ou, semelhante ao inglês <i>rat</i> , mas com o dorso da língua mais para trás. Fricativa pós-alveolar vozeada, ou aproximante pós-alveolar vozeada.
<i>z</i>	/ts/	Como no japonês <i>tsunami</i> . Não como “zebra”. Sibilante alveolar não-vozeada africana.
<i>c</i>	/tʃʰ/	Semelhante ao <i>z</i> acima, mas aspirado (acompanhado de sopro de ar). Sibilante alveolar não-vozeada africana aspirada.
<i>s</i>	/s/	
<i>w</i>	/w/	Como em “uai”. Não é pronunciada antes de ⟨u⟩ e ⟨ü⟩.
<i>y</i>	/j/	Como em “iate”. Não é pronunciada antes de ⟨i⟩. Antes de final ⟨u⟩, denota o som /y/, como na final ⟨ü⟩ (q.v.).

### Finais

<i>a</i>	[a]	
<i>e</i>	[ɛ]	Semelhante ao /o/ português, mas com os lábios como em /e/; semelhante ao inglês <i>duh</i> , mas mais próximo de <i>mom</i> . Vogal semi-fechada posterior não-arredondada.
<i>ao</i>	[aʊ]	
<i>an</i>	[an]	Com o /n/ articulado de fato, com a ponta da língua.
<i>en</i>	[ən]	Como no inglês <i>coven</i> ; vogal semelhante à interjeição “dã”, seguida de /n/ articulado com a ponta da língua.
<i>ang</i>	[aŋ]	Semelhante a <i>an</i> acima, mas com o <i>zpy-ng</i> articulado com o fundo da língua.
<i>eng</i>	[ɛŋ]	Semelhante a <i>en</i> acima, mas com o <i>-ng</i> articulado com o fundo da língua.
<i>ong</i>	[oŋ]	Vogal “u” leve, como o final de “favo” na pronúncia brasileira, ou o inglês <i>book</i> .

<i>er</i>	[ɚ]	Vogal neutra como em <i>en</i> acima. Consoante como o “r” do inglês americano ou português caipira.
<i>i</i>	[i]	Geralmente como no português; mas, depois de <i>z, c, s, zh, ch, sh, r</i> , apenas continua a consoante, sem ser pronunciada como vogal (.
<i>iao</i>	[iəʊ]	Como no português “miau”.
<i>iu</i>	[iʊ]	Note o [o] não-grafado.
<i>ian</i>	[iɛn], [iæn]	Vogal nuclear como no “é” português, ou então como no inglês <i>cat</i> .
<i>iong</i>	[iʊŋ]	Vogal não é um o mas um “u” leve, como o final de “favo” na pronúncia brasileira, ou o inglês <i>book</i> .
<i>u</i>	[u]	
<i>uo, o</i>	[ʊo]	Escrito sem ⟨u⟩ depois de ⟨b,p,m,f⟩.
<i>ui</i>	[ʊeɪ]	Note o [e] não-grafado.
<i>un</i>	[ʊən]	Vogal nuclear neutra, como no inglês <i>won</i> ; semelhante ao português “dã”.
<i>ü</i>	[y]	Como no ⟨ü⟩ alemão ou ⟨u⟩ francês. Como o /i/ português, mas arredondando os lábios como em um /u/.