

3

Apresentação da Língua e Escrita Chinesa

Para o estrangeiro que procure estreitar contato com a civilização chinesa, a língua daquele país parece uma barreira quase insuperável. Contribuem para isso o exotismo dos sons e a dificuldade em adquirir e identificar os tons, além da enorme homofonia da língua. Num primeiro contato, o falante não nativo se pergunta como lidar com uma língua aparentemente monossilábica que tem apenas cerca de 400 sílabas básicas¹⁶⁴ para a expressão de todos os significados naquela língua? E por cima de tudo, há a enorme tarefa de aprender uma escrita que desafia o estudante na sua complexidade e aparente falta de regularidade. Assim, alguns métodos de ensino privilegiam o estudo do chinês apenas utilizando-se da sua transliteração oficial, o *pīnyīn* (拼音, literalmente: “juntando os sons”), uma ortografia basicamente fonêmica e que se utiliza de sinais diacríticos acima das vogais em cada sílaba para indicar os tons. É talvez a única língua não ágrafa do mundo hoje, a qual na prática, um falante que a tenha como L2 possua um acesso e possa se comunicar (verbalmente) com os falantes nativos sem que tenha qualquer domínio de sua escrita.¹⁶⁵ Várias gramáticas da língua para estrangeiros são inclusive escritas apenas utilizando-se do *pīnyīn*.¹⁶⁶ Poderíamos até argumentar que nesta prática está sendo usado um sistema real de escrita, que o estudante poderá manter seu caderno de notas, fazer exercícios e praticar leituras, tudo usando apenas o *pīnyīn*. Não é absolutamente um

¹⁶⁴ Teoricamente são até 1.200 combinações silábicas se considerarmos os fonemas suprasegmentais (tons), que entretanto normalmente são deixados à margem pelo estudante iniciante do chinês.

¹⁶⁵ É claro que com o uso da transliteração de outras línguas (como o árabe ou hindi) para um alfabeto ocidental é também possível que o estudante que queira usar essas línguas na comunicação oral não precise aprender o seu sistema de escrita. Todavia, por se tratar de sistemas de grafemas muito mais simples que o chinês, isso na prática acontece muito raramente e só no estudo mais superficial.

¹⁶⁶ Dois exemplos já usados pelo próprio autor são Li & Thompson (1981) e Po-Ching & Rimmington (1997).

aprendizado ágrafo, simplesmente estaria sendo empregado um sistema de escrita não nativo.¹⁶⁷

Entretanto não parece haver dúvidas que caso o estudo de uma língua tenha por motivação a procura do conhecimento daquela língua e daquele povo e uma experiência de alteridade, ao invés da mera preparação para uma negociação comercial ou uma rápida viagem de turismo, por exemplo, a língua chinesa não pode absolutamente ser dissociada de sua escrita. Talvez não se possa nem mesmo dizer estritamente que está se aprendendo chinês ou usando chinês “plenamente” sem que se esteja travando contato com a sua escrita, que é tão particular e intrinsecamente ligada aos modos de expressão do chinês.

Acredito que a fala chinesa já foi tão influenciada por sua escrita – bem como a influenciou – que não podemos falar de uma sem pensar na outra. Em suma, que não há língua chinesa em sua plenitude sem escrita chinesa.¹⁶⁸ A escrita chinesa não é de forma alguma mero reflexo da sua fala, pelo passivo da utilização de uma linguagem para fins comunicativos, descritivos ou expressivos. Sem a escrita chinesa não haveria a China, e muitos discursos, principalmente aqueles que se afastam mais da uma linha estritamente representacionista, parecem apoiar esta afirmação. Espero que isso se torne mais claro ao final do presente trabalho.

3.1. Breve Introdução à Língua Chinesa

Não é o objetivo deste trabalho uma discussão exaustiva sobre a língua chinesa, mas sim uma pesquisa sobre os estudos em que a escrita chinesa se articula no conjunto da língua e da civilização chinesa. Entretanto, para o

¹⁶⁷ Sistemas de escrita não nativos tiveram uso extenso no processo de educação e de aprendizado de língua ágrafas, como em línguas indígenas. Todavia parece que é apenas no caso do chinês em que se utiliza tão extensivamente uma notação estrangeira padronizada (alfabética) para uma língua que já tem uma tradição de escrita própria.

¹⁶⁸ Isso não significa obviamente que o analfabeto chinês não possa usar a língua como uma eficiente ferramenta expressiva e comunicativa, mas sim que o conhecimento e uso pleno da língua acarreta o saber escrever e que a própria língua chinesa em seu desenvolvimento e autoconhecimento não deve ser dissociada da escrita.

benefício daqueles que têm pouca familiaridade com a língua chinesa, neste capítulo será feita uma breve apresentação, com especial foco naqueles aspectos que ajudarão a melhor compreender as discussões posteriores específicas sobre a escrita sinítica.

Os fatos aqui apresentados são basicamente incontroversos e, quando não o forem, será feita alguma observação específica. As fontes de informação são basicamente os dicionários e os livros sobre chinês, além das gramáticas que constam da bibliografia. Outras referências serão citadas diretamente ao longo do texto.

3.1.1. Apresentação

A língua chinesa pertence a um ramo independente (sinítico) da família sino-tibetana de línguas e é falada por cerca de 91% dos habitantes da República Popular da China,¹⁶⁹ além de ser a língua oficial de Taiwan e uma das quatro línguas oficiais de Cingapura.

Depois do Indo-europeu (com estimados 2,0 bilhões de falantes), o Sino-tibetano é o segundo ramo linguístico com mais falantes do mundo, cerca de 1,04 bilhões, a imensa maioria composta pelos falantes de chinês mandarim, com cerca de 700 milhões, além de 250 milhões de falantes dos “dialetos” do mandarim, 22 milhões de falantes do birmanês, língua oficial de Mianmá e 4 milhões de falantes do tibetano. Há mais de 300 outras línguas relacionadas à este ramo, todas com uma distribuição demográfica muito restrita.¹⁷⁰

Sua distribuição geográfica se estende basicamente pela China, Tibete e partes de Mianmá.

Uma diferenciação importante que devemos fazer diz respeito à polissemia do termo “chinês,” dedicado à língua mas também ao povo de etnia Han (*hàn*, 汉

¹⁶⁹ O mandarim é falado por cerca de 70% da população da China, a diferença (em relação aos 91%) corresponde aos outros “dialetos” ou “regionaletos” do chinês.

¹⁷⁰ Para maiores detalhes, veja-se: Comrie, 1990, capítulo 40.

/ 漢), que corresponde a cerca de 93% (censo de 1982) da população da China, e também até mesmo a todos os habitantes do país, que formalmente são “chineses” (*zhōngguó rén*, 中国人, literalmente: “pessoas do país do meio”). A etnia Han deve seu nome à grande dinastia Han que unificou o império e reinou sobre a maior parte do que hoje é o país, entre 206 aC e 220 dC. As demais etnias correspondem a cinquenta e cinco minorias étnicas formalmente reconhecidas (*shǎoshù mǐnzú*, 少数民族, “minorias de população étnica”), tais como os mongóis (蒙古人, *měnggú rén*), tibetanos (西藏人, *xīzàng rén*), os uigures (维吾尔人, *wéiwúěr rén*) e os Hui (回族人, *húizú rén*). Essas minorias falam línguas não-chinesas, exceto pelos Hui, que é a denominação dos cerca de 10 milhões de muçulmanos falantes do chinês.

O mandarim falado tem algumas características marcantes que se mostrarão especialmente importantes na sua articulação com a escrita:

- A escrita é a única em uso hoje em que basicamente se utiliza de um significante monossilábico passível de atribuição de significado. Embora no chinês clássico isso significasse a forte tendência a palavras monossilábicas, no mandarim contemporâneo há uma abundância de palavras polissilábicas. Existem raros exemplos de caracteres que (no chinês atual) nunca são empregados de forma isolada, como por exemplo os dois caracteres de 蟋蟀, *xīshuài*, que juntos significam “gafanhoto” e não são usados separadamente.¹⁷¹
- O mandarim tem uma relativa pobreza fonológica (como veremos abaixo). Isso, somado à ligação próxima entre morfemas e sílabas provoca uma enorme homofonia na língua. Num exemplo comumente utilizado, um site de aprendizado de chinês/inglês na Internet¹⁷² indica nada menos do que 178 significados para a sílaba “ji” em *pīnyīn*. Há uma certa compensação por se tratar de uma língua tonal, com quatro tons (alto, ascendente, alto-baixo e descendente) no mandarim padrão. Em relação a esses tons, o fenômeno de *sândi tonal* é complexo e os tons mudam quando em contato uns com os outros por uma questão de eufonia e facilidade de pronúncia.
- O chinês é uma língua isolante ou analítica (em oposição às línguas consideradas sintéticas e às aglutinantes): As regras de formação de palavras por aposição (sufixação, prefixação ou circunfixação) em chinês são relativamente pouco produtivas e não há declinação gramatical.

¹⁷¹ Tecnicamente os dois caracteres significam “gafanhoto,” seriam sinônimos e poderiam ser usados isoladamente com este significado, mas na prática é um dissilábico. Todavia a questão do significado dos caracteres que compõem esses dissílabos é questionável, como veremos mais adiante neste trabalho.

¹⁷² www.yellowbridge.com, acesso em Setembro/2010.

- O chinês utiliza-se predominantemente de uma estrutura tópico/objeto, em oposição à relação sujeito/predicado, mais comumente usada pelas gramáticas das línguas ditas “ocidentais.”
- O chinês é considerada uma língua orientada pelo contexto no sentido em que ela não requer a codificação de certas formas gramaticais que não acrescentem novas informações (neste sentido ela é mais “econômica” ao descartar estruturas formais gramaticais que não acrescentem informações).
- Em chinês não há flexão verbal indicativa de tempo (igualmente não há flexões para indicar número, gênero ou modo). A ação é situada no tempo basicamente através de advérbios e de outros elementos indicados nas frases ou pelo próprio contexto. A própria língua opta por dar ênfase na demonstração sintática do aspecto da ação (por exemplo, ação completada ou habitual).
- Da maneira similar a qual substantivos coletivos em português têm especificadores especiais, como em “uma *matilha* de cães” ou “uma *alcatéia* de lobos,” também os substantivos em chinês têm seus *especificadores*. Todavia, todos os substantivos devem ser classificados (às vezes de maneira idiossincrática) em grandes grupos através desses especificadores. Às vezes os especificadores podem até ser usados para substituir o substantivo.¹⁷³

Nas seções a seguir serão tratados com um pouco mais de detalhes aspectos fonéticos/fonológicos e morfológicos do chinês moderno, importantes para este trabalho.

Para fins deste trabalho estarei usando intercambiavelmente os termos *chinês*, *mandarim* ou *pǔtōnghuà* (普通话, literalmente “língua comum”) para se referir sempre ao regionalismo (“dialetto”) padrão da China continental e falado na região de Beijing.

3.1.2. Fonética e Fonologia

A forma das sílabas em chinês é bastante restrita: não há combinações consonantais (como “fr” em “fran” de frango ou “cl” em “clas” de classe) e a sílaba chinesa (no *pǔtōnghuà*) é restrita ao seguinte modelo:

(C)V(V)(V ou N), sendo:

¹⁷³ Novamente há similaridade no caso dos coletivos em português. Por exemplo: “O caçador foi perseguido o dia inteiro pela *alcatéia* faminta.”

() – parênteses mostram opção

C – consoante (inicial)

V – vogal

N – consoante nasal (/n/ e /ŋ/)

Há dois tipos de fonemas em chinês: os fonemas segmentais, que podem ser pensados como os sons sequenciais (tais como /b/, /ŋ/, /a/, etc.) e os fonemas suprasegmentais, ou tons, que de certa forma são “adicionados” às sílabas como um todo e formam pares mínimos fonológicos.

Ainda assim a sílaba chinesa é mais complexa do que a japonesa, por exemplo: o japonês tem apenas 113 sílabas diferentes, enquanto o chinês tem 1.277 sílabas se considerarmos os tons, e um número entre 398 e 418 “sílabas segmentais” ou “sílabas básicas” se desconsiderarmos os fonemas suprasegmentais (tons).¹⁷⁴ Já o inglês, no outro extremo, tem potencialmente mais de 15.000¹⁷⁵ sílabas diferentes ao passo que uma sílaba em inglês pode conter até 7 fonemas (como, por exemplo, em “splints”), comparadas as 4 do chinês.

3.1.3. Morfologia e Léxico

O que distingue um substantivo de um verbo ou adjetivo em chinês não é a declinação (que não existe), mas a existência de uma *palavra especificadora*, um “contador de unidades” daquele substantivo. E estritamente não há a classe gramatical de adjetivos em chinês, já que eles se comportam como verbos. O mais correto seria considerar a qualificação de adjetivos como verbos adjetivais.

¹⁷⁴ Jorden, 1963:xxi, in DeFrancis, 1996, p.42.

¹⁷⁵ Um estudo de Chris Barker, professor do departamento de linguística da NYU calculou um total de 15.831 sílabas diferentes em inglês (<http://semanticsarchive.net/barker/Syllables/index.txt>, acesso em 11/02/2011).

Embora existam alguns sufixos indicadores de nome, em especial 子, *zi*, a maior parte dos substantivos em chinês não se distingue de outras classes gramaticais. Por exemplo, 死, *sǐ*, pode indicar “morte”, “morto” ou mesmo a locução adverbial “até a morte” dependendo do contexto e de outras palavras.¹⁷⁶

Há uma tendência à composição lexical entre os ideogramas. Por exemplo, as palavras quase sinônimas como *chàng* e *gē*, (唱 e 歌), ambas com a acepção de “cantar,” “canto,” “música,” são colocadas juntas pela tendência de cunho fonológico - do chinês contemporâneo - de formar “palavras” dissilábicas. Ambos os grafemas são independentes, mas sintaticamente o composto *chànggē* (唱歌) pode ser usado com se fosse uma única palavra. Esta situação torna ainda mais complexa a definição de *palavra* em chinês, como veremos mais a frente neste trabalho.

Em contraste com a aposição, o processo de composição em mandarim tem uma grande produtividade. Estima-se que cerca de 2/3 do vocabulário cotidiano de Beijing constitua-se de polissilábicas, esmagadoramente dissilábicas.¹⁷⁷ Além do tipo de composição citada acima, outros processos ocorrem em mandarim: compostos verbais resultativos, compostos verbais paralelos, compostos nominais, compostos nome-verbo, reduplicação e os casos bem restritos de afixação.

3.2. Apresentando a Escrita Chinesa

Neste capítulo será feita uma apresentação dos principais aspectos da escrita chinesa, tal como vistos através dos diversos trabalhos que estudaram esse sistema de escrita.

¹⁷⁶ Esta é uma característica também comum (embora não tão difundida quanto em chinês) do inglês, em que um grande número de verbos pode ser usada como substantivo, por exemplo “call”, “play”, “talk”, “walk”, etc

¹⁷⁷ Comrie, 1990, p.817.

A escrita chinesa tem um óbvio aspecto gráfico bidimensional. Embora em um primeiro momento ela pareça ser um conjunto impossivelmente complexo de centenas ou mesmo milhares de pequenos “desenhos” diferentes entre si, o sistema de escrita chinês (*hànzì*, 汉字, “caracteres chineses”) é dotado de uma complexa estrutura que organiza espacialmente seus componentes gráficos. Conquanto diversos autores tenham uma leitura bem diferente acerca da forma como estes caracteres são usados, do seu impacto cognitivo, dificuldade de uso, etc., os padrões que sublinham o sistema de escrita chinês são basicamente incontestes.

Os caracteres são “construídos” a partir de um repertório limitado da ordem de uma dezena de *traços*, que irão se combinar em caracteres simples indecomponíveis (exceto nos traços que o formam), da ordem de algumas centenas. Tais *caracteres simples* (*dútǐzì*, 独体字, literalmente “caracteres de corpo único”) poderão ser usados em pareamentos de dois ou mais para formar *caracteres complexos* (*hétǐzì*, 合体字, literalmente “caracteres de corpo juntado”). O número de traços nos caracteres simples e complexos varia muito e há caracteres simples com mais traços do que complexos (por exemplo, *miàn*, 面, “face”, “lado” tem 9 traços e é simples ao passo que *hǎo*, 好, tem 6 traços e é composto de *nǚ*, 女 e *zǐ*, 子). O caractere mais simples é aquele que indica a unidade e o número 1: *yī*, 一. Por outro lado, há disputas em relação a escolha do caractere mais complexo e alguns deles dão mostra da imensa criatividade dos calígrafos chineses na combinação de caracteres simples. Sem dúvida um dos mais complexos é aquele usado para um tipo de macarrão da região de Shaanxi, cuja pronúncia não padrão é *biáng* e que compõem-se de 57 traços:¹⁷⁸

¹⁷⁸ Para todos efeitos práticos na escrita usualmente se utilizam substitutos fonéticos mais simples do que este, como 棒棒麵 (*bàng bàng miàn*) ou 梆梆麵 (*bāng bāng miàn*), sendo que 麵 é o caractere tradicional para *macarrão* e *massas em geral*. DeFrancis (1984, p.75) indica um caractere com 64 traços, *tiè*, significando “verborrágico.”

Independente do número de traços e da complexidade do caractere, todos eles devem ocupar um “quadrado imaginário” de tamanho constante, o que traz harmonia e beleza ao texto chinês, da mesma forma como exigiu um complexo trabalho de programação gráfica na constituição de alguns caracteres.

3.2.1. Tipos de Caracteres

Tradicionalmente os caracteres chineses são classificados nas seis categorias de Xu Shen, introduzidas no seu livro *shuōwén jiězì* (说文解字, “Explicação dos caracteres simples e análise dos caracteres compostos”), publicado há quase 2000 anos. As categorias de Xu Shen são:

- 象形, *xiàngxín*, pictogramas, literalmente “aparece na forma”

São os caracteres que surgem em toda apresentação sobre a língua chinesa e que ilustram o “pitoresco” lado pictográfico da escrita chinesa. Exemplos típicos são:

- | | |
|-----------------------------|---|
| ○ 山, <i>shān</i> , montanha | 牛, <i>niú</i> , boi |
| ○ 雨, <i>yǔ</i> , chuva | 木, <i>mù</i> , árvore |
| ○ 人, <i>rén</i> , pessoa | 本, <i>běn</i> , raiz (derivado de “árvore”) |
| ○ 门, <i>mén</i> , porta | 口, <i>kǒu</i> , boca. |

Podemos observar que mesmo nas ocasiões em que o caráter pictográfico torna-se claro após a sua explicação, na maior parte das vezes (senão em todas) não seria possível adivinhar o significado simplesmente olhando o caractere isoladamente. A estilização progressiva da escrita chinesa também relegou ao passado algumas claras indicações pictográficas, como a de 月, *yuè*, “lua.”

- 指示, *zhǐshì*, caracteres indicativos, literalmente “indicar e mostrar”

- 假借, *jiǎjiè*, caracteres emprestados, literalmente “emprestar e tomar emprestado”

São caracteres empregados de uma forma adaptada (ou não) para formar um segundo caractere, normalmente emprestando seu aspecto fonético. Assim 长, *cháng*, “longo” é usado de forma gráfica idêntica em 长, *zhǎng*, “líder” ou 令 *lìng*, “ordem” é usando para compor 领 *lǐng*, “liderar”. Se assemelha ao processo do rébus sem o acréscimo de um indicador semântico (que formaria um *xíngshēng*).

Estudiosos contemporâneos procuraram adaptar a classificação original de Xu Shen para o chinês de hoje, normalmente através de alguma combinação das 6 categorias originais. Não cabe aqui uma apresentação detalhada das diversas classificações propostas, mas sim o que hoje é aceito pela maioria dos paleógrafos chineses, as três categorias propostas por Chen Mengjia em 1956.¹⁷⁹

- Pictógrafos: a categoria é auto-explicativa. É uma mistura dos *pictogramas* e *caracteres indicativos* de Xu Shen e contém elementos que pictoricamente aludam à sua significação, ainda que esta ligação seja muito tênue, especialmente à medida que os caracteres foram evoluindo graficamente ao longo dos séculos.
- Grafemas emprestados (*loan graphs*): um caractere usado inicialmente para designar um objeto ou ideia é “emprestado” para outra significação. O uso atual de diversos sinais em mensagens de texto por celular ou email exemplifica este processo (“rébus”) na língua portuguesa, como por exemplo: “me encontre @ casa”, o símbolo “@” substituindo a ideia da preposição “em”. Assim 来, *lái*, era originalmente um pictógrafo de uma planta de trigo e foi posteriormente usado para designar “vir” (entre outras acepções), que tinha esta pronúncia na língua falada. 能, *néng*, era o pictógrafo de um tipo de urso, hoje significa “poder, ter a capacidade de”, ao passo que 熊, *xióng*, ainda hoje usa-se para designar “urso.” Os empréstimos por vezes se apoiavam em semelhanças fonológicas e o mesmo

¹⁷⁹ Veja-se Kane, 2006, p.32-34.

princípio foi muito utilizado na simplificação da escrita chinesa adotada pela China continental, como no exemplo de 后, *hòu* que hoje significa tanto “imperatriz” (后, *hòu*, não simplificado) como “depois” (后, *hòu*, não simplificado).

- Gráficos fonéticos: a grande maioria dos caracteres atuais: compostos de duas partes, um *radical* (quase todos pictógrafos), que dá uma indicação semântica ao caractere e um componente fonético, que indica a pronúncia. Se assemelha ao *xíngshēng* e ao *jiǎjiè* de Xu Shen.

A tipologia dos caracteres será importante nas discussões posteriores sobre as características semânticas e fonéticas da escrita chinesa. Uma das discussões mais polêmicas sobre a escrita chinesa, veremos, lida exatamente com a representatividade e importância dos caracteres acima chamados de *gráficos fonéticos*.

3.2.2. Traços

Os caracteres chineses são escritos com traços. Esses traços são de três tipos: pontos, linhas e “ganchos” (linhas com uma espécie de gancho na ponta). Por sua vez os traços são subdivididos de acordo com a sua orientação (por exemplo: linha do alto à esquerda para baixo à direita). Os traços poderiam ser pensados em correspondência gráfica à 2ª articulação de Martinet, aquela desprovida de significado.

Uma lista num livro sobre caligrafia chinesa¹⁸⁰ aponta 24 traços, ao passo que um dicionário chinês-inglês¹⁸¹ indica apenas 12 traços. A diferença entre estas avaliações basicamente é derivada de um maior rigor na separação de traços muito

¹⁸⁰ Fazzioli e Mei Ling, 1986, p.20.

¹⁸¹ *Oxford Beginner's Chinese Dictionary*, 2000.

semelhantes. Apesar de tais discrepâncias, não há ambiguidade na identificação do que se constitui um traço.

Não só o repertório de traços é bastante regulado, como também a ordenação em que os traços devem ser escritos segue uma série de princípios gerais que devem ser obedecidos. Considera-se errado um caractere que, mesmo que tenha o desenho final escrito correspondente à norma, não tenha sido escrito com os traços na ordem certa. As regras mais gerais são: 1) os traços são escritos da esquerda para a direita;¹⁸² 2) são escritos de cima para baixo; 3) os traços horizontais precedem os verticais. Estas três regras valeriam para 90% dos caracteres em uso corrente.¹⁸³

Já vimos que o número de traços em um caractere pode variar de um até mais de 30. Com a simplificação dos caracteres ocorrida na China continental, o número médio de traços dos 2.000 caracteres mais usuais caiu de 11,2 para 9,8.¹⁸⁴

Os traços podem ser considerados como constituindo o nível mais baixo (mais simples) da hierarquia sobre a qual é construída a notação da escrita chinesa, nomeadamente: traços → caracteres simples → caracteres compostos.

3.2.3. Radicais

Os gráficos que remetem a alguma ideia ou noção semântica são extremamente importantes para o chinês, tanto o clássico como o contemporâneo, embora, como veremos, o chinês antigo tenha tido um percentual muito maior de caracteres de base unicamente semântica. O conceito de *radical* é chave no entendimento deste tipo de caractere/gráfico.

¹⁸² Mesmo no tempo em que a direção geral da escrita chinesa era ambígua, os caracteres já eram escritos da esquerda para a direita.

¹⁸³ DeFrancis, 1984, p.77. Para mais detalhes, veja-se *Oxford Beginner's Chinese Dictionary*, 2000, p.165-166.

¹⁸⁴ Ping, 1999, p.157. No mesmo livro, p.136, Ping indica um trabalho (Wu e Ma, 1988, *Hanzi he hanzi gaige shi*. Changsha: Hunan Renmin Chubanshe) que estima 11 traços em média para os 7.000 caracteres mais usuais

O radical deve oferecer alguma informação semântica ao caractere em questão. Alguns são quase óbvios, como por exemplo o radical 口, *kǒu*, que pictograficamente assemelha-se a uma boca, usado em caracteres que remetem a conceitos relacionados à fala e à boca e também àqueles caracteres que simplesmente refletem algum tipo de som: 叫, *jiào*, “chamar”; 听, *tīng*, “ouvir”; 吃, *chī*, “comer”; 啥, *shá*, “falar o que pensa”; 嗓, *sǎng*, “garganta/voz”; 吸, *xī*, “inalar/absorver”; 唱, *chàng*, “cantar”; 喊, *hǎn*, “berrar”; 吹, *chuī*, “soprar”; 叹, *tàn*, “suspirar” e as partículas “sonoras”: 吧, *ba*; 呢, *ne*; 吗, *ma*; 啦, *la*; 呀, *ya*; 啊, *a*; e assim por diante. A língua chinesa é pródiga em listas como essa, mas não podemos ficar com a impressão de uma simplicidade que é bastante ilusória. O radical 口 é um daqueles que mais claramente classificam um grupo de compostos e pode dar uma ideia equivocada da utilidade e eficiência dos radicais. Para cada exemplo deste tipo há dezenas de outros bem mais “misteriosos,” como veremos.

Nos exemplos com 口 acima fica óbvio identificar o que é o radical e o que é a parte fonética do caractere. Há todavia caracteres muito mais complexos em que essa identificação não só não é simples, como por vezes não há consenso entre os estudiosos da língua sobre o que exatamente constitui o radical. Um exemplo é 牙齿, *yáchǐ*, “dente” em chinês. Ambos os caracteres isoladamente também significam “dente” e o segundo caractere é um caractere-radical, ou seja, funciona isoladamente e também como indicador semântico em outros caracteres (por exemplo, 龄, *líng*, idade/período de tempo e 龋, *qǔ*, cárie).¹⁸⁵ Mesmo um caractere muito comum como 出, *chū*, “sair” por vezes aparece classificado sob o

¹⁸⁵ Exemplos de idiossincrasias na classificação de caracteres abundam no chinês. Por exemplo, embora 冬, *dōng*, “inverno” use um radical de água simplificado (os dois traços abaixo do caractere), ele é atualmente classificado sob o radical 攴, *zhī*, “seguir.”

radical 凵, *qū*, “receptáculo” enquanto outras fontes indicam o radical 丨, *gǔn*, “número 1,” “linha.”¹⁸⁶

Em consonância com tamanha irregularidade, naturalmente não poderia haver um número de radicais que sempre tenha sido fixo e determinado. O *shuōwén jiězǐ* (说文解字) escrito por Xu Shen tinha 540 radicais. O dicionário *kāngxī* (康熙字典), publicado em 1716 adotou um sistema mais simples baseado num dicionário editado 100 anos antes e listou 214 radicais. Para todos efeitos o dicionário Kangxi tornou-se a referência padrão no chinês atual e todos os radicais são geralmente referidos como *Kangxi X*, sendo X o número do radical na listagem Kangxi. No exemplo acima 𪗇 é Kangxi 211. Apesar da padronização do Kangxi, outros esquemas são adotados em importantes dicionários hoje em dia, nomeadamente os 187 radicais *Xīnhuā* (新化) usado no *xīnhuā zìdiǎn* (新化字典) e os 155 radicais usados no importante dicionário *xiàndài hànyǔ cídiǎn* (现代汉语词典, “dicionário do chinês contemporâneo”)

A seguir a lista dos radicais Kangxi:

¹⁸⁶ Num caso ilustrativo apresentado por DeFrancis (1984, p.93), um dicionário popular entre estudantes publicado nos EUA em 1945 listava 7.773 caracteres sob o título de “Lista de Caracteres com Radicais Obscuros,” referência que era de grande ajuda ao leitor nos tempos pré-informáticos.

214 KangXi 康熙字典 部首/Bushou/Radicals Scheme

One Stroke									
1 一	2	3 丶	4 丿	5 乙	6 丿				
Two Strokes									
7 二	8 亅	9 人	10 儿	11 入	12 八	13 冂			
14 冫	15 冫	16 儿	17 冫	18 冫	19 冫	20 冫			
21 匕	22 匕	23 匕	24 十	25 卜	26 巾				
28 厶	29 又								
Three Strokes									
30 口	31 口	32 土	33 士	34 又	35 宀	36 夕			
37 大	38 女	39 子	40 山	41 寸	42 小	43 巾			
44 千	45 彡	46 山	47 厶	48 工	49 己	50 巾			
51 干	52 彡	53 厶	54 厶	55 井	56 彡	57 巾			
58 彡	59 彡	60 彡							
Four Strokes									
61 心	62 戈	63 户	64 手	65 支	66 支	67 文			
68 斗	69 斤	70 方	71 无	72 日	73 母	74 比			
75 木	76 欠	77 止	78 夕	79 夕	80 母	81 比			
82 毛	83 氏	84 气	85 水	86 火	87 爪	88 文			
89 文	90 彡	91 片	92 牙	93 牛	94 犬				
Five Strokes									
95 玄	96 玉	97 瓜	98 瓦	99 甘	100 生	101 用			
102 田	103 疋	104 牙	105 石	106 目	107 皮	108 血			
109 目	110 彡	111 矢	112 石	113 示	114 肉	115 禾			
116 穴	117 立								
Six Strokes									
118 竹	119 米	120 糸	121 缶	122 网	123 羊	124 羽			
127 老	126 而	127 来	128 耳	129 革	130 肉	131 良			
132 白	133 至	134 白	135 舌	136 舛	136 舛	137 衣			
138 色	140 艸	141 疋	142 虫	143 血	144 舟	145 衣			
146 西									
Seven Strokes									
147 見	148 角	149 言	150 谷	151 豆	152 豕	153 豕			
153 辰	155 赤	156 走	157 足	158 身	159 里	160 辛			
161 辰	162 疋	163 邑	164 邑	165 菜	166 菜				
Eight Strokes									
167 金	168 長	169 門	170 阜	171 隶	172 隹	173 雨			
174 青	175 非								
Nine Strokes									
176 面	177 羊	178 耑	178 韭	180 音	181 頁	182 風			
182 飛	184 食	185 首	186 香						
Ten Strokes									
187 馬	188 骨	189 高	190 彭	191 鬥	192 鬯	193 高			
194 鬼									
Eleven Strokes									
195 魚	196 鳥	197 鹵	198 鹿	199 麥	200 麻				
Twelve Strokes									
201 黃	202 黍	203 黑	204 蒲						
Thirteen Strokes									
205 龍	206 鼎	207 鼓	208 鼠						
Fourteen Strokes									
209 鼻	210 齊								
Fifteen Strokes									
211 齒									
Sixteen Strokes									
212 龍	213 龜								
Seventeen Strokes									
214 禽									

© Dylan WH Sung 2000

Seja qual foi a listagem usada, os radicais são sempre ordenados por contagem de traços, desde um (一, 乙, entre outros) até (no caso da classificação Kang Xi) os 17 de 龠, *yuè*, que significa “instrumento musical.” Um radical pode ter uma forma gráfica “livre” e outra forma derivada quando funciona como radical. O exemplo mais comum é de 水, *shuǐ*, “água,” que aparece como 4 “gotas” (氵) na forma de um radical, por exemplo, de 江 e 深, respectivamente *jiāng*, “rio” e *shēn*, “profundo.”¹⁸⁷ Em outro exemplo, 言, *yán*, que significa “palavra,” “falar,” “linguagem” e era usado no mesmo formato (讠) como radical (por exemplo, em 說, *shuō*, “falar”), na escrita simplificada da China continental foi modificado para 讠 (no mesmo exemplo, 说, *shuō*, “falar”).

Já vimos que via de regra o radical imprime uma conotação semântica ao caractere, porém existem muitos deles que são pouco mais do que um identificador gráfico para indexar caracteres. Podemos citar Coulmas (2003) quando lista alguns dos significados mais idiossincráticos entre os radicais chineses: “cabeça de porco,” “sombra,” “veneno,” “melão,” “cão,”¹⁸⁸ ao que eu poderia acrescentar “veado” (radical KangXi 198), “um recipiente raso” (KangXi 108), “dar uma leve batida/pancadinha” (KangXi 66), etc. As idiossincrasias não param por aí. Peguemos por exemplo novamente o radical de água. Ele gera caracteres com uma clara ligação com os conceitos de “água,” “molhado,” “mar”, etc. Como explicar entretanto 治, *zhì*, “governo,” ou 派, *pài*, “enviar”, “despachar”, “grupo”, “facção,”¹⁸⁹ ou 沽, *gū*, “comprar e vender”? Claro, dentre as centenas de caracteres com o radical de “água,” a grande maioria tem algum tipo de ligação com *água*, mas estamos falando de um dos radicais mais óbvios e

¹⁸⁷ O dicionário eletrônico *ABC Chinese-English Comprehensive Dictionary* indica 924 caracteres com este radical.

¹⁸⁸ Coulmas, 2003, p.56.

¹⁸⁹ Na verdade poderíamos explicar diacronicamente que o caractere para 派 possuía o significado original de “rio tributário”, e que por um processo de rébus foi utilizado com as acepções acima. Este fato porém não elimina sua inutilidade sincrônica como um indicador semântico.

produtivos, e mesmo assim existem exceções, e, mais ainda, com caracteres de uso muito comum. E essa é a norma na escrita chinesa.

Vimos portanto como não há nada parecido com uma taxonomia de conceitos ou sistematização.¹⁹⁰

3.2.4. O Repertório de Indicadores Fonéticos

Ao contrário dos radicais que apontam para o eixo semântico, aqueles indicadores fonéticos que compõe a 2ª parte dos caracteres *xíngshēng* não possuem uma classificação tão elaborada.

Dois motivos certamente são: 1) variações fonéticas enormes dentro dos regionaletos da língua chinesa, resultado do descasamento normal no processo diacrônico de mudança da fala (mais rápido) do que da escrita (mais lento) e; 2) grande diversidade e número destes indicadores, resultado de um processo diacrônico muito produtivo na criação de compostos semântico/fonéticos. Os radicais neste sentido são muito mais adequados às necessidades classificatórias dos chineses. Por outro lado, o interesse dos linguistas pela relação entre a escrita chinesa e a fala suscitou diversas tentativas no sentido de procurar classificar ou pelo menos identificar quais e quantos são os indicadores fonéticos na escrita chinesa. DeFrancis (1984, p.93-115) descreve uma análise extensa das características fonéticas da escrita chinesa, que serão apresentadas aqui e, posteriormente trazida de volta à discussão ao longo deste trabalho.

A primeira lista foi publicada por um missionário católico, J. M. Callery, que em 1841 listou um silabário de 1.040 caracteres. Quarenta anos mais tarde W.W. Soothill classificou 4.300 caracteres em 895 indicadores fonéticos.¹⁹¹ Outras listas bem mais extensas foram propostas, assim como análises daqueles caracteres mais prolíficos na criação de compostos *xíngshēng*. Ping cita, sem

¹⁹⁰ Veja-se DeFrancis, 1984, p.93.

¹⁹¹ Ao passo que há uma análise autóctone chinesa extensa sobre a classificação dos radicais, a classificação dos indicadores fonéticos parece ter surgido no trabalho de sinólogos ocidentais. Esse fato parece reforçar o olhar chinês mais voltado para os aspectos semânticos de sua escrita.

referências externas, “cerca de 1.300 símbolos distintos funcionando como determinantes fonéticos” (Ping, 1999, p.141). Coulmas indica que “vários dicionários listam entre 888 e 1040 indicadores fonéticos” (Coulmas, 2003, p.57).

As listas propostas variam enormemente entre cerca de 1.000 até 6.000 indicadores fonéticos e refletem uma tentativa de impor uma ordem na massa extremamente heterogênea dos caracteres chineses, submetida a séculos de decisões classificatórias e pressões linguísticas. Entretanto, DeFrancis indica recentes estudos convergindo para uma faixa mais estreita entre 900 e 1300 caracteres (mesmo que tal classificação jamais tenha atingido o *status* oficial dos radicais).

Ao contrário dos radicais, há uma forma de se avaliar a funcionalidade e eficiência dos indicadores fonéticos. Isso porque a fonologia das línguas chinesas (como a de todas as outras) tem um repertório de número aproximado e limitado de fonemas e de combinações silábicas que deveriam corresponder de uma forma ou outra aos indicadores fonéticos. Mais ainda, é possível comparar a escrita chinesa com outros sistemas de escrita, também na relação entre seus grafemas e os fonemas/sílabas da língua. Desta forma poderiam ser criadas algumas estatísticas para uma avaliação mais quantitativa da representatividade fonética da escrita chinesa frente a outros sistemas de escrita.¹⁹²

Alguns autores, como Kane, chegam a praticamente rejeitar os indicadores fonéticos como minimamente eficientes na sua função:

Mudanças históricas (...) significaram que o sistema dos indicadores fonéticos, mesmo com os caracteres tradicionais, era tão irregular que na prática se mostrou totalmente enganoso. Fazia mais sentido considerar os indicadores fonéticos como a parte residual dos caracteres, além do radical.¹⁹³

Kane, entretanto, não oferece maiores justificativas para tal afirmação além de uma lista óbvia de várias inconsistências e idiosincrasias. Outros autores, entretanto, fizeram alguns levantamentos de *corpus* que podem ajudar a avaliar melhor a situação da indicação fonética da escrita chinesa.

Por exemplo, das cerca de 1.277 sílabas tonais possíveis em mandarim, apenas 558 aparecem na tentativa de silabário de Soothill feita na década de 1880.

¹⁹² A conclusão de tais análises, embora aproxime (ou não) escrita e fala (em sua foneticidade) não implica necessariamente em considerar a escrita como instrumento de representação da fala. Veremos esta discussão em mais detalhes adiante neste trabalho.

¹⁹³ Kane, 2006, p.50.

Estas sílabas estão representadas por 895 indicadores fonéticos. Muitos deles apresentam-se em uma relação 1:1 com as sílabas e aparentemente é a característica tonal das sílabas que se representaria de forma menos consistente. Alguns casos, como da sílaba *yì*, chegam a ter 12 grafemas diferentes. Comparativamente, um estudo em inglês mostra que seus cerca de 40 e poucos fonemas (*forty-odd sounds*) são ortografados em mais de 600 formas diferentes, grafemas isolados ou conjuntos deles.¹⁹⁴ Veja que Soothill não tentou fazer uma classificação fonética científica da língua e escrita chinesa, apenas mostrar quais seriam os indicadores fonéticos num corpus de caracteres usuais.

Nos diversos processos de criação de caracteres, como vimos acima, parece não ter havido qualquer padronização no uso dos indicadores fonéticos, dependendo apenas da vontade pessoal e dos padrões estético-visuais de seus criadores, bem como da aceitação dos usuários e do poder normativo do poder central chinês. Mais ainda, a evolução gráfica dos caracteres, embora lenta, sem dúvida afetou o que poderia já ter sido no passado uma melhor “taxa de representatividade fonética,” problema agravado na China continental pela simplificação dos caracteres ocorrida no século XX. DeFrancis (1984, p.101-105) analisa em detalhes a tentativa de Soothill em demonstrar como este sistema, apesar de todas as influências diacrônicas, ainda oferece uma indicação muito melhor do que aquela através dos radicais (os indicadores semânticos).

Uma análise importante sobre a eficácia dos indicadores fonéticos é apresentada numa amostra feita por DeFrancis de 500 caracteres de um corpus feito por Chen Heqin¹⁹⁵ de 900.000 caracteres contendo 4.719 caracteres diferentes, ordenados por frequência dentro do corpus. Seu resultado indica que nos 100 caracteres mais frequentes 48% eram caracteres com um composto fonético identificado por Soothill, mas somente em 37% destes casos a indicação sugeriu corretamente a sua pronúncia.¹⁹⁶ Entretanto este percentual de representatividade rapidamente aumenta para os caracteres menos frequentes e atinge 66% em todos os 500 caracteres amostrados por DeFrancis, sendo deste percentual: 24,6% com identidade completa; 17% com identidade completa,

¹⁹⁴ Zachrisson R.E. (1931), *Four hundred years of English spelling reform*, citado em DeFrancis (1984, p.97).

¹⁹⁵ Chen Heqin (1928), citado em DeFrancis, 1984, p.107.

¹⁹⁶ Neste corpus, os 100 caracteres mais comuns totalizam 47% do total de realização dos 900.000 caracteres.

exceto tom; e 24,4% com identificação de pelo menos um dos fonemas segmentais do caractere composto. Estes percentuais parecem confirmados em Ping (1999): de uma lista de 5.990 caracteres *xíngshēng* comuns, 26,3% são pronunciados exatamente igual ao determinante fonético,¹⁹⁷ só que este autor, que concede à fonética um papel bem menos importante do que DeFrancis, indica que pelo menos 20% dos caracteres em uso comum no chinês moderno não tem qualquer indicação fonética.¹⁹⁸

O estudo acima, a despeito de ser um importante indicador sobre as características fonéticas da escrita chinesa, foi feito com base em dados da primeira metade do século XX, com caracteres tradicionais (não simplificados). DeFrancis também apresenta outro trabalho feito em 1978 com um corpus de 8.075 caracteres, dos quais 98% são compostos fonéticos ou caracteres simples fonéticos. Destes, 3.117 (48%) são considerados como possuindo indicadores fonéticos úteis (idênticos em pronúncia, exceto pelos tons). Somados ao elementos fonéticos simples (17%) chegamos a 65% dos caracteres que são foneticamente identificáveis.¹⁹⁹

Finalmente, um outro estudo de 1988 indica que de uma lista publicada em 1952 dos 2.000 caracteres mais usuais, 74% seriam compostos fonéticos.²⁰⁰

Os dados apresentados acima estão resumidos na tabela a seguir:

Tabela 4- Grau de foneticidade nos caracteres chineses

Núm caract.	Carac. fonét. simples	Compostos fonéticos	Indicadores fonéticos válidos nos compostos	Fonte
100	52%	48%	37%, sendo 10% c/correspond. exata	DeFrancis (1984), p.106, estudo 1928
500 ²⁰¹	18%	79%	66%, sendo 25% c/correspond. exata	
2.000		74%		Ping (1999), p.135, estudo 1952
5.990			26,3% c/correspond. exata	Ping (1999), p.142, estudo 1992
8.075	17%	81%	48%	DeFrancis (1984), p.109 (corpus 1978)

¹⁹⁷ Ping, 1999, p.142.

¹⁹⁸ X. Qiu (1988), *in* Ping, 1999, p.142.

¹⁹⁹ DeFrancis, 1984, p.109-110.

²⁰⁰ Qiu Xigui (1988). *Wenzixue gaiyao*. Beijing: *Shangwu Yinshuguan*, citado em Ping, 1999, p.135.

²⁰¹ Amostra entre os 4100 caracteres mais frequentes do corpus de Chen Heqin.

3.2.5. Número de Caracteres

A apresentação diacrônica da escrita chinesa irá mostrar como o repertório de caracteres chineses cresceu enormemente ao longo de sua história. O grande número de radicais e indicadores fonéticos, somado ao fato de que não há limite teórico para o número de caracteres simples que podem ser combinados em um composto torna o número potencial de caracteres virtualmente astronômico. Por outro lado, já desde o dicionário Kang Xi do século XVIII (que já contava com 48 mil caracteres) que o crescimento desse número parece estar seguindo uma linha assintótica, que tenderia aos 60 mil caracteres. O Estado chinês comunista, num espírito de padronização da escrita chinesa, hoje opõe-se abertamente à criação de novos caracteres (o que ocorre então de forma mais livre em lugares de maior liberdade política e criativa, como Hong Kong e Taiwan), com a notável exceção dos termos de elementos químicos.²⁰²

De qualquer forma, trata-se de um número altíssimo, certamente uma impossibilidade prática além dos limites da memória humana, até mesmo porque sua organização interna, como vimos acima, está longe de ser regular.

Ajudaria na avaliação da complexidade lexical do chinês estudar quais são, na prática, as reais necessidades de conhecimento e uso dos caracteres. Veja-se que, no corpus preparado por Chen Heqin em 1928 e citado acima, textos somando 900.000 caracteres apresentaram apenas 4.719 caracteres diferentes. Um outro exemplo muito citado, pela notoriedade de seu autor, é o *Collected Works of Mao Zedong (1917-49)*, que inclui também cerca de 900 mil caracteres escritos e apenas 3.136 deles únicos.²⁰³

A consulta dos caracteres raros e pouco usuais sempre poderá ser feita através de dicionários e sistemas informáticos, da mesma forma que o falante do inglês tem a sua disposição uma referência no *Oxford English Dictionary*, ainda

²⁰² Para uma discussão mais detalhada, veja-se Allenton, 2008, p.128-130.

²⁰³ Kane, 2006, p.53. Todavia Rogers, 2005, p.44, cita 2.981 caracteres únicos para a mesma coletânea.

mais rico lexicalmente, já que sua versão de 2005 contava com cerca de 301 mil entradas.²⁰⁴

Uma relação de cobertura de caracteres num corpus levantado de publicações recentes na China fornece os seguintes percentuais²⁰⁵:

Núm. caracteres	Cobertura
500	80%
1.000	91%
2.400	99%
3.800	99,9%
5.200	99,99%
6.600	99,999%

O *List of Common Characters in Modern Chinese* foi publicado na China continental em Janeiro de 1998 pela Comissão Estatal para a Linguagem. A Parte I contém os 2.500 caracteres mais usados e a Parte II contém 1.000 caracteres “menos usados.” A Lista de Caracteres Regulares em Chinês Moderno de Março de 1988, publicada pela Comissão Estatal para a Linguagem e Editora do Estado, contém 7.000 entradas.²⁰⁶

O limite mínimo em geral considerado para julgar uma pessoa “funcionalmente alfabetizado” é de 2.000 caracteres²⁰⁷ e espera-se que estudantes da escola primária tenham aprendido cerca de 2.500 caracteres, passando para 3.500 caracteres ao tornarem-se estudantes universitários. Viviane Allenton indica o conhecimento de 5.000 caracteres para tais estudantes e a mesma autora estima 3.500 caracteres frequentes, 3.000 com “frequência inferior” e mais 5.000 para uso especializado.²⁰⁸ Rogers fala em 1.500 para alfabetização básica, 2.500-3.000 para alfabetização funcional e 6.000 para estudiosos.²⁰⁹

Finalmente, o teste oficial do governo chinês continental para estudantes de chinês como L2, chamado de HSK (*hànyǔ shuǐpíng kǎoshì*, 汉语水平考试, “teste

²⁰⁴ Fonte: Wikipedia, “Oxford English Dictionary,” consultado em 02/09/10.

²⁰⁵ Dados de Chang (1989) *Xiandai Hanyu pinlü cidian de yanzhì*; Su (1992) *Xiandai hanzi yanjiu jianshu* e Yin (1991) *Xiandai hanzi de dingliang yanjiu*. Citados em Ping, 1999, p.137.

²⁰⁶ Ping, 1999, p.136.

²⁰⁷ Ibid., p.136.

²⁰⁸ Allenton, 2008, p.50.

²⁰⁹ Rogers, 2005, p.44.

de nivelamento da língua chinesa”) apresenta claramente quantos caracteres e palavras são exigidos para cada um dos seus 4 níveis:²¹⁰

- Nível de base: 798 caracteres, 1.033 palavras
- Nível elementário-intermediário: 1.606 caracteres, 3.051 palavras
- Nível intermediário: 2.215 caracteres, 5.253 palavras
- Nível avançado: 2.915 caracteres, 8.822 palavras.

Há portanto um número estimado entre 2.500 e 5.000 caracteres cujo conhecimento é necessário para o uso da escrita chinesa em quase todas as esferas não especializadas, ou seja, para considerar que seu usuário seja plenamente alfabetizado.

²¹⁰ Poizat-Xie, 2007, p.4. Recentemente os testes padronizados do HSK foram modificados e atualmente existem 6 níveis ao invés dos quatro aqui mostrados.