Convenções usadas na conversão Grafema-Fonema (G2P) com o alfabeto SAMPA

Autores:
Fernando Perdigão, Sara Candeias, Arlindo Veiga

Instituto de Telecomunicações – Polo de Coimbra
Coimbra, abril de 2011

1. Introdução

Este documento descreve as convenções que foram usadas no sistema de conversão entre grafemas e fonemas quanto aos símbolos admitidos, tanto para fonemas quanto para grafemas.

O alinhamento entre grafemas e fonemas fica simplificado se forem usados símbolos com um só caráter para fonemas, como por exemplo /õ/ em vez de /o~/, pois evita a inserção do grafema correspondente ao til. De forma idêntica, os dígrafos correspondem a um só fonema, como é o exemplo de <ss> que dá origem ao fonema /s/. Evita-se neste caso o apagamento de um grafema.

Verificámos que a necessidade da inserção de fonemas é particularmente reduzida. De forma a evitar-se completamente a inserção de fonemas, foram definidos símbolos para mais que um fonema (Tabela 2).

As tabelas seguintes convencionam os símbolos usados do conjunto ISO-8859-1[1], bem como exemplos de utilização com o alfabeto SAMPA[2].

Apresentam-se também algumas notas sobre o problema do alinhamento entre grafemas e fonemas e a tabela de "custos" usada no programa de alinhamento.

Optámos por nos referir a fonemas e não a fones, uma vez que o estudo é fundamentalmente fonológico, sem esmiuçar a sua realização acústica em alofones.

[1] http://pt.wikipedia.org/wiki/ISO_8859-1

[2] http://www.phon.ucl.ac.uk/home/sampa/

Tabela 1 - Alfabeto SAMPA e extensões com um só caráter para alinhamento G2P.

nº	Símbolo SAMPA	Uni-caráter (SAMPAuc)	Grafemas possíveis	Exemplo/Observação
Vogais				
1	6	6 | ª	a, e, â	cama, senha, câmara
2	a		a, á, à	pá, pala
3	@		e	de
4	e		e, ê	vê, dedo
5	E		e, é,	pé, pele
6	i		i, í, y, e	vi, aí, henry, reunião
7	o		o, ô, ou	oco, avô, louco
8	O		o, ó	pó, pote
9	u		u, ú, o	tu, tio, to, baú
10	6~	ã	ã, an, em, am, êm,	branco
11	e~	ë	ên, en, em	pente, agência, empate
12	i~	ï	i, in, im, ím, ín, m	sim
13	o~	õ	õ, ôn, on, om	ponte, cônsul
14	u~	ü	u, ún, un, um	atum
Semivogais (não usadas)				
15	j			
16	w			
17	j~	ì		
18	w~	ù		
Plosivas/Oclusivas				
19	b		b	belo
20	d		d	dado
21	g		g, gu	gato
22	p		p	pato
23	t		t	tu
24	k		qu, c, k	casa, aquela,
Fricativas				
25	f		f	fé
26	s		s, cc, cç, ç, x, c+{eiéí}	sol, céu, trouxe, caça
27	S		ch, s, z, x	chave, xá, paz, pás
28	v		v	viu
29	z		z, s, x	casa, zé, exemplo
30	Z		j, g, s, z, x	já, gira, desviar, ex-bar
Líquidas				
31	l		l	lá
32	L		lh	velho
33	r		r	caro
34	R		r, rr	carro, rato
Consoantes nasais				
35	m		m	mão
36	n		n	nada
37	J		nh	senha

Tabela 2 - Vogais em posição tónica e símbolos para múltiplos fonemas

nº	Símbolo SAMPA	UniCaráter (SAMPAuc)	Grafemas possíveis	Exemplo/Observação
Vogais em posição tónica				
38	"6	â		cama → /k"am6/ → /kâm6/
39	"a	á		casa → /k"az6/ → /káz6/
40	"e	ê		tema → /t"em6/ → /têm6/
41	"E	É		sete → /s"Et@/ → /sÉt@/
42	"i	í		tio → /tíu/
43	"o	ô		ovo → /ôvu/
44	"O	Ó		logo → /l"Ogu/ → /lÓgu/
45	"u	ú		uva → /úv6/
46	"6~	Ã		campo → /kÃpu/
47	"e~	Ë		centro → /s"e~tru/→/sËtru/
48	"i~	Ï		cinco →/s"i~ku/→/sÏku/
49	"o~	Õ		conto → /k"o~tu/ → /kÕtu/
50	"u~	Ü		assunto → /6s"u~tu/→/asÜtu/
Símbolos para múltiplos fonemas				
51	6i	æ	e+{xj}	extrair → /æStr6ír/
52	"6i	Æ	e+{xj}	extra →/ ÆStr6/;
53	6~i~6~	Ê	{tv}+êm	têm → /t"6~i~6~i~/ → /tÊï/
54	o~i~	¤	p**õ**em:	/po~i~6~i~/→ /p¤ãï/
55	ks	K	ficção, fax	
56	ai	Å	caem	/k"ai6~i~/→/kÅãï/
57	Oi	®	constroem	/ko~StrOi6~i~/→/kÖStr®ãï/
Fonemas Alternativos/Adicionais/Opcionais (para uso futuro em modelos acústicos)				
58	i|e	y	**e**sófago, **e**leitor	/iz"Of6gu/ vs /ez"Of6gu/
59	o|O	Ô	**o**lhar	/oL"ar/ vs /OL"ar/
60	e|E	È	**e**ólica, **e**tnia, h**e**braica	/E"Olik6/ vs /e"Olik6/
61	u|ø	µ	q**u**ente	
62	e|6	å		
63	l~	£		
64	ø	_	h, -,' '	(fonema nulo)
65	-	0	-	(marcador de silêncio)
66	#	#	'sp', ' '	(sp; separador de palavras)

Nota: para as semivogais não são usados os símbolos /j/, /w/, /j~/ e /w~/. Optámos por notá-las como /i/, /u/, /i~/ e /u~/, respetivamente.

Tabela 3 - Grafemas especiais para dígrafos

grafema	dígrafo/letras	fonema(s)	Exemplo com grafemas especiais
C	cc	s, ks	ficcional → fiCional
Ç	cç	s, ks	ficção → fiÇão
R	rr	R	carro → caRo
§	ss	S	massa →ma§a
L	lh	L	molho → moLo
J	nh	J	unha → uJa
S	ch	S	chave → Save
º	ou	o	dourada → dºrada
Ä	an, am	ã, Ã	canto →cÄto, campo →cÄpo
Ë	en, em	ë, Ë	sente →sËte, sempre → sËpre
Ï	in, im	ï, Ï	limbo→lÏbo
Ö	on, om	õ, Õ	conto→cÖto, dom→dÖ
Ü	un, um	ü, Ü	assunto, um
Â	ân, âm	Ã	pântano →pÂtano, lâmpada → lÂpada
Ê	ên, êm	Ë	ênfase → Êfase, êmbolo →Êbolo
Í	ín, ím	Ï	índio→Ídio, límpido→lÍpido
Ô	ôn, ôm	Õ	cônsul→cÔsul, cômputo→cÔputo
Ú	ún, úm	Ü	denúncia→denÚcia, cúmplice→cÚplice

2. Notas sobre alinhamento entre grafemas e fonemas

Princípios:

1. O alinhamento entre grafemas e fonemas permite fazer uma associação ótima entre símbolos, podendo existir apagamentos e inserções de grafemas e fonemas. Exemplos desta associação podem ser os seguintes:

```
<e_xtra> <põ_em> <tê__m> <fix_o> <campo> <milho> <guerra/
/6jStr6/ /põïãï/ /tãïãï/ /fiksu/ /kã_pu/ /miL_u/ /g_ER_6/
```

2. A definição dos símbolos para dígrafos e de símbolos para fonemas múltiplos tem como objetivo simplificar a associação entre grafemas e fonemas de tal forma que o alinhamento seja, tanto quanto possível, unívoco. Desta forma, é possível fazer com que não existam inserções de grafemas. Apenas em alguns casos existe inserção de fonemas (por exemplo no caso das formas nasaladas com n ou m: <an/→/ã_/).
Evita-se a inserção de fonemas (ou apagamento de grafemas) considerando 4 símbolos que agregam mais que um fonema:

K = /ks/;	exemplo: <ficção> →/fiÇão/→/fiKãü/; fixar →/fiKar/
æ = /6i/;	exemplo: <e_xtra> →/6iStr6/ vs <extra/→/æStr6/
Æ = /6~i~6~/=/ãïã/;	exemplo: <tê__m> →/tãïãï/ vs <têm/→ /tÆï/
¤ = /o~i~/=/õï/;	exemplo: <põ_em> →/põïãï/ vs <põem/→/p¤ãï/

3. Os custos de apagamento e de inserção de grafemas e/ou de fonemas, bem como os custos de associação entre grafemas e fonemas são previamente calculados. Dá-se preferência à inserção e ao apagamento de fonemas face a uma substituição incorreta. Por exemplo, nos alinhamentos seguintes, apenas o último é aceitável. Isto consegue-se afetando um custo elevado às associações r/@, e/S, n/d e S/@ no 1ª caso e s/e no 2º caso, baixando, em contrapartida, o custo da inserção dos grafemas r, n, s (ou apagamento dos fonemas correspondentes).

```
<correspondesse> <correspondesse> <correspondesse/
/kuR@S_põde_s@_/ /kuR_@Spõ_d_es@/ /kuR_@Spõ_des_@/
```

Casos mais complicados são os seguintes:

```
<trans_eunte> <t_ens>
/trã_ziü__t@/ /tãï_S/
```

pois o grafema <u> pode ser inserido (ex: g**u**erra) assim como o <n> (em dígrafos nasais), e o <e> pode corresponder a /ï/ (ex: <ões> → /õïS/). O fonema /ã/ também pode ser facilmente inserido, como se viu antes. A solução passa por baixar o custo das associações: <u>→ /ü/, <e> → /i/ e <n> → /ĩ/.
(ver alignG2P.m)

Apêndice: Matriz de custos

```
char Rabc[] = "aáãâàbcçdeéêèëfghiíìjklmnoóõôòpqrstuúûùvwxyzCÇR§LJSºÄËÏÖÜÂÊÍÔÚ _-0"; //grafemas
char Tabc[] = "6a@eEioOuãëïõüjwìùbdgptkfsSvzZlLrRmnJâáêÉíôÓúÃËÏÕÜæÆÊ¤KyÔÈµå£_0#";   //fonemas
char delTable[66]=\
//a á ã â à b c ç d e é ê è ë f g h i í ì j k l m n o ó õ ô ò p q r s t u ú û ù v w x y z C Ç R § L J S º Ä Ë Ï Ö Ü Â Ê Í Ô Ú   _ - 0
{ 8,8,8,8,8,8,6,8,8,8,8,8,8,8,8,8,6,8,8,8,8,8,8,8,5,5,8,8,8,8,8,8,8,5,5,8,5,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,5,8,5,8 };
char insTable[64]=\
//6 a @ e E i o O u ã ë ï õ ü j w ì ù b d g p t k f s S v z Z l L r R m n J â á ê É í ô Ó ú Ã Ë Ï Õ Ü æ Æ Ê ¤ K y Ô È µ å £ _ 0 #
{ 8,8,8,8,8,5,8,8,8,5,8,5,8,8,8,8,8,8,8,8,6,8,8,8,8,5,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8, };
char ConfusionTable[66][64]={\
//6 a @ e E i o O u ã ë ï õ ü j w ì ù b d g p t k f s S v z Z l L r R m n J â á ê É í ô Ó ú Ã Ë Ï Õ Ü æ Æ Ê ¤ K y Ô È µ å £ _ 0 #
{ 1,6,9,9,9,9,9,9,9,4,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,7,3,9,9,9,9,9,9,6,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, // 1:a -> 6,a,ã,â,á,Ã
{ 9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, // 2:á -> a,á
{ 9,9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, // 3:ã -> ã,Ã
{ 2,9,9,9,9,9,9,9,9,3,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,2,9,9,9,9,9,9,9,3,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, // 4:â -> 6,ã,â,Ã
{ 9,2,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, // 5:à -> a,á
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, // 6:b -> b
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,1,9,2,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,7,9,9,9,9,9,9,5,9,9 }, // 7:c -> k,s,K,_
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, // 8:ç -> s
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, // 9:d -> d
{ 6,9,1,6,5,6,9,9,9,5,4,6,9,9,9,9,6,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,6,9,5,5,9,9,9,9,5,5,9,9,9,7,7,9,9,9,7,9,7,9,6,9,7,9,9 }, //10:e -> 6,@,e,E,i,ã,ë,ï,ì,â,ê,É,Ã,Ë,æ,Æ,y,È,å,_
{ 9,9,9,9,2,9,9,9,9,6,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,2,9,9,9,9,6,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //11:é -> E,ã,É,Ã
{ 7,9,9,3,9,9,9,9,9,7,2,7,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,7,9,3,9,9,9,9,9,7,2,9,9,9,7,7,7,9,9,9,9,9,9,9,9,9,9,9 }, //12:ê -> 6,e,ã,ë,ï,â,ê,Ã,Ë,æ,Æ,Ê
{ 9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //13:è -> E,É
{ 9,9,9,9,7,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //14:ë -> E,É
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //15:f -> f
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,2,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //16:g -> g,Z
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9 }, //17:h -> _
{ 9,9,9,9,9,2,9,9,9,9,9,5,9,9,3,9,3,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,3,9,9,9,9,9,7,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //18:i -> i,ï,j,ì,í,Ï
{ 9,9,9,9,9,2,9,9,9,9,9,6,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,2,9,9,9,9,9,6,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //19:í -> i,ï,í,Ï
{ 9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //20:ì -> i
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //21:j -> Z
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //22:k -> k
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,2,5,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,2,9,9,9 }, //23:l -> l,L,£
{ 9,9,9,9,9,9,9,9,9,9,9,4,9,3,9,9,4,3,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,2,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,4,9,9 }, //24:m -> ï,ü,ì,ù,m,_
{ 9,9,9,9,9,9,9,9,9,9,9,7,9,9,9,9,7,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,2,5,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,1,9,9 }, //25:n -> ï,ì,n,J,_
{ 9,9,9,9,9,9,7,5,1,9,9,9,4,5,9,9,9,5,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,4,5,7,9,9,9,7,9,9,9,9,9,9,9,7,9,9,9,9,9,9,9 }, //26:o -> o,O,u,õ,ü,ù,ô,Ó,ú,Õ,Ô
{ 9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //27:ó -> O,Ó
{ 9,9,9,9,9,9,9,9,9,9,9,9,2,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,2,9,9,9,9,7,9,9,9,9,9,9,9,9,9,9 }, //28:õ -> õ,Õ,¤
{ 9,9,9,9,9,9,2,9,9,9,9,9,4,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,2,9,9,9,9,9,4,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //29:ô -> o,õ,ô,Õ
{ 9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //30:ò -> O,Ó
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,7,9,9 }, //31:p -> p,_
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //32:q -> k
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,1,4,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,5,9,9 }, //33:r -> r,R,_
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,3,1,9,4,6,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,5,9,9 }, //34:s -> s,S,z,Z,_
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //35:t -> t
{ 9,9,9,9,9,9,9,9,2,9,9,9,9,6,9,4,9,4,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,4,9,9,9,9,7,9,9,9,9,9,9,9,9,3,9,9,3,9,9 }, //36:u -> u,ü,w,ù,ú,Ü,µ,_
{ 9,9,9,9,9,9,9,9,2,9,9,9,9,5,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,2,9,9,9,9,5,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //37:ú -> u,ü,ú,Ü
{ 9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //38:û -> u
{ 9,9,9,9,9,9,9,9,7,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //39:ù -> u,ú
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //40:v -> v
{ 9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //41:w -> u,v
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,5,1,9,3,6,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,3,9,9,9,9,9,9,9,9,9 }, //42:x -> s,S,z,Z,K
{ 9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //43:y -> i
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,4,9,1,7,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,7,9,9 }, //44:z -> S,z,Z,_
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,4,9,9,9,9,9,9,9,9,9 }, //45:C -> s,K
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,4,9,9,9,9,9,9,9,9,9 }, //46:Ç -> s,K
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //47:R -> R
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //48:§ -> s
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //49:L -> L
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //50:J -> J
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //51:S -> S
```

```
{ 9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //52:º -> o
{ 9,9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,2,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //53:Ä -> ã,Ã
{ 9,9,9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,2,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //54:Ë -> ë,Ë
{ 9,9,9,9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,3,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //55:Ï -> ï,Ï
{ 9,9,9,9,9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,4,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //56:Ö -> õ,Õ
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,1,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,2,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //57:Ü -> ü,Ü
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //58:Â -> Ã
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //59:Ê -> Ë
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //60:Í -> Ï
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //61:Ô -> Õ
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //62:Ú -> Ü
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,2,2,2 }, //63:  -> _,0,#
{ 9,9,9,9,9,2,9,9,9,3,9,3,9,9,9,9,9,9,9,9,6,9,9,9,9,3,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9 }, //64:_ -> i,ã,ï,g,s
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,2,2,2 }, //65:- -> _,0,#
{ 9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,9,0,9 }, //66:0 -> 0
//6 a @ e E i o O u ã ë ï õ ü j w ì ù b d g p t k f s S v z Z l L r R m n J â á ê É í ô Ó ú Ã Ë Ï Õ Ü æ Æ Ê ¤ K y Ô È µ å £ _ 0 #
//| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
//a a e e e e o o o a e e o m i u e m b d g p t c f c s v s g l l r r m n n a a e e i o o o a e i o u e e ê õ c e o e u e l c
//â á   ê é i ô ó u ã ê ê õ o     i o     _     k   ç x w x j   L   R     J â á ê é í ô ó u ã ê í õ ú ê ê     x             e - -
//e à     è í º ò ú â Ë i ô u     m u           q   s z   z s               e à   è     ò ú â Ë Ï ô Ü         C             h 0
//ê       ë ì     û e   í Ö ú     n                 x S     x               ê     ë       ù e Ê Í Ö Ú         Ç             m
//          y     ù é   m   Ü                       C       z                               é     Ô                         n
//          _     w ê   n                           Ç                                       ê                               p
//                  Ä   Ï                           §                                       Ä                               r
//                  _   _                           _                                       Â                               s
//                                                                                                                          u
//                                                                                                                          z
//
//                                                                                                                          -
};
```