

TR-S-0005

A design of phonetic-balanced transcript for Chinese

宇佐美慶 張勁松
Kei Usami Jin-Song Zhang

2000年9月8日

Abstract

This technical report presents our efforts to develop phonetic-balanced transcripts for Chinese, which are supposed necessary for building a Chinese speech data corpus. The search algorithm was based on maximum entropy rule by which an equal appearance of the phonetic context is preferred. As the preliminary experimental results, we found a 7-utterance-set in which nearly all Initials and Finals appear, and a 25-utterance-set in which 134 di-phones appear. This system is supposed to be used to generate more specifically phonetic-balanced transcripts for Chinese in the future.

実習報告書

テーマ :

中国語音声データベース用
文セットの設計

実習先 : (株) エー・ティ・アール

音声言語通信研究所 第一研究室

担当者 : Jin-song ZHANG (客員研究員)

実習生 : 宇佐美 慶 (龍谷大学)

目次

1. 目的 1
2. 理論 1
2. 1 音韻バランス	
2. 2 エントロピー	
2. 3 中国語音韻	
2. 3. 1 音節の構造	
2. 3. 2 Initialsと子音	
2. 3. 3 Finals	
2. 4 文セット抽出アルゴリズム	
3. 実験内容と結果 5
3. 1 monosyllableでの実験	
3. 2 Initial+Final での実験	
3. 3 di-phoneでの実験	
4. 考察 10
5. 最後に 10

1. 目的

音声認識、音声合成、音声品質等の研究用の文音声データベースを構築するためには、その基本となるデータベース(文セット)が必要である。また、文音声データベースはできるだけ少ない数の文で、音声現象をよく反映しているものが望ましい。そこで今回、音韻の出現頻度を考慮した音韻バランスのとれた文セットを作成する方法として、情報エントロピーを評価尺度としたプログラムをつくり、長・短の文が入った母集団から文セットを抽出する事を目的とした。

2. 理論

2.1 音韻バランス

音韻のバランスが取れている状態は、音韻の出現頻度が等しい状態と考えることができる。この状態は、音韻を確立事象とした場合のエントロピーが最大の状態であると考えられることができる。

2.2 エントロピー

エントロピーは次式で定義される。

$$(式1) \quad S = - \sum_{n=1}^N p(n) \log p(n)$$

ここで $p(n)$ は文セット中に現れる音素環境 n ($n=1 \sim N$ 、 $\sum p(n)=1$)の相対出現頻度である。

2.3 中国語音韻

2.3.1 音節の構造

中国語の音韻は、音節を「Initial」と「Final」に分けたものである。「Initial」は、子音で始まる音節であり、「Final」は「Initial」の音節を除いたものである。ここで、ta, qi, jin, chuangを例にとると、それぞれの「Initial」は t-, q-, j-, ch-となり、それぞれの「Final」は -a, -i, -in, -uangとなる。また、「Final」はGlide、Kernel、endingの3つに分かれている。「Final」-uangの -uはGlide、-aはKernel、-ngはending codaというように分けられる。

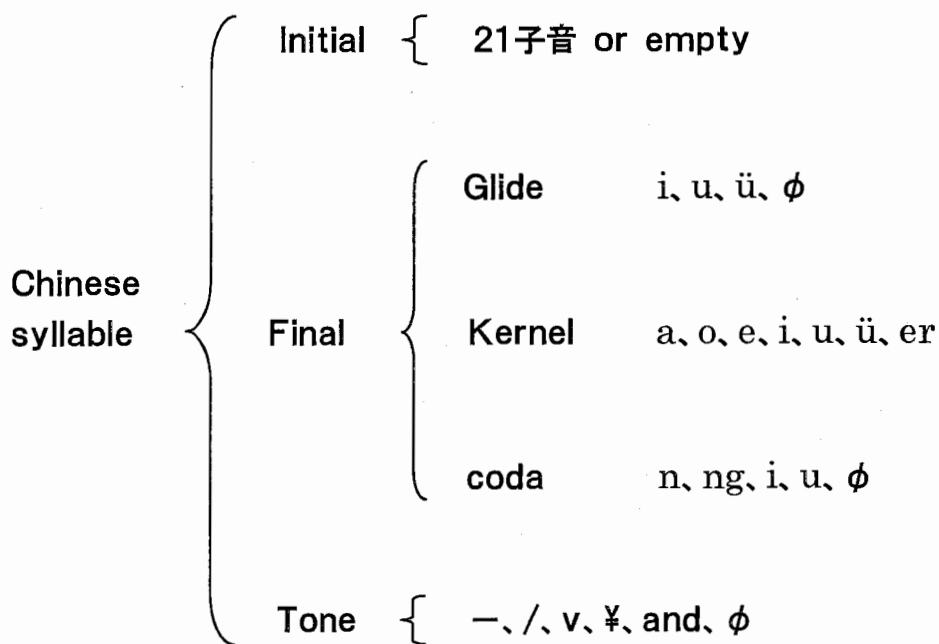


図:1 中国語音節の構造

2.3.2 Initialsと子音

一般的な中国語の子音は22種類ある。それを 表:1 に示した。鼻子音の中の/ng/を除く21種類の子音は「Initial」に、唯一/ng/だけは音節のendingに入る。

2.3.3 Finals

音韻の「Final」の種類は39ある。それを 表:2 に示した。

Place	Manner	Stops(閉鎖音)		Affricative(破擦音)		Fricative(摩擦音)		Nasal (鼻音)	Lateral (側音)
		Unasprt	Asprt	Unasprt	Asprt	Unvcd	Vcd		
Labial(唇音)		b	p			f		m	
Dentals(歯音)		d	t					n	l
Dental Sibilants(歯擦音)				z	c	s			
Retroflexes(そり舌音)				zh	ch	sh	r		
Palatals(口蓋音)				j	q	x			
Gutturals(喉音)		g	k			h		ng	

表:1 明瞭な発音の位置づけによる中国語子音の区分け

Compound Complexities	Glide-based Final Classification			
	Row-a	Row-i	Row-u	Row-ü
Single vowels and Open ending Finals	-ĭ	i	u	ü
	a	ia	ua	
	o		uo	
	e			
	ê	ie		üe
Compound Vowels	er			
	ai		uai	
	ei		uei	
	ao	iao		
Nasal Ending Compound	ou	iou		
	an	ian	uan	üan
	en	in	uen	ün
	ang	iang	uang	
	eng	ing	ueng	
		ong	iong	

表:2 中国語音節の「Final」s

Category	Finals
1	ĭ, ĩ
2	i, ia, ie, iai, iao, iou, ian, in, iang, ing
3	u, ua, uo, uai, uei, uan, uen, uang, ueng
4	ü, üe, üan, ün, üng
5	a, ai, ao, an, ang
6	e, ei, ê, en, eng, er
7	o, ou

表:3 Glideにより7分類された39種の「Final」s

2.4 文セット抽出アルゴリズム

文セットの文数を M 、母集団の文数を N とする。

[1] 母集団の上から M 文を抽出して仮の文セット V とし、残りの $N-M$ 文の集合を U とする。

[2] 仮の文セット V から1文($V(i)$)を抽出する。 ($1 \leq i \leq M$)

[2.1] 仮の文セット V のエントロピー P を計算する。

[2.2] 文集合 U から1文($U(t)$)を抽出する。 ($1 \leq t \leq (N-M)$)

[2.2.1] $V(i)$ と $U(t)$ を交換したときの仮の文セット V のエントロピー $S(t)$ を計算する。 ($S(1) \sim S(N-M)$)

[2.2.2] $t=t+1$ 進んで、[2.2]へもどる。

もし、 $t=(N-M)+1$ の時[2.3]へ。

[2.3] $S(1)$ から $S(N-M)$ の中で一番大きいエントロピー $S(k)$ が P より大きければ $V(i)$ と $U(k)$ を交換する。

[3] $i=i+1$ 進んで、[2]へもどる。

もし、 $i=M+1$ の時、もう一度 $i=1$ から[2]を繰り返す、一度も $V(i)$ と $U(k)$ の交換がなければ終了。

3. 実験内容と結果

今回、monosyllable、Initial+Final、di-phone の3パターンの音韻と母集団を用意し、それぞれの文セットの抽出を行った。

3.1 monosyllableでの実験

ここで用意した音韻は、411種類である。

母集団は22220文あり、仮の文セットを200文とした。

最初の仮の文セット200文のエントロピーは6.9889であった。そしてこの仮の文セットの音韻の出現頻度にばらつきがあり、音韻バランスがわるかった。

ここで、文セット抽出処理を行った

最終的な結果

文セットのエントロピーは8.4503になった。いくつか出現しなかった音韻はあったが、音韻バランスの取れた文セットを抽出することができた。この最適化過程におけるエントロピーの変化を 図:2 に表わした。

最適化過程におけるエントロピー変化

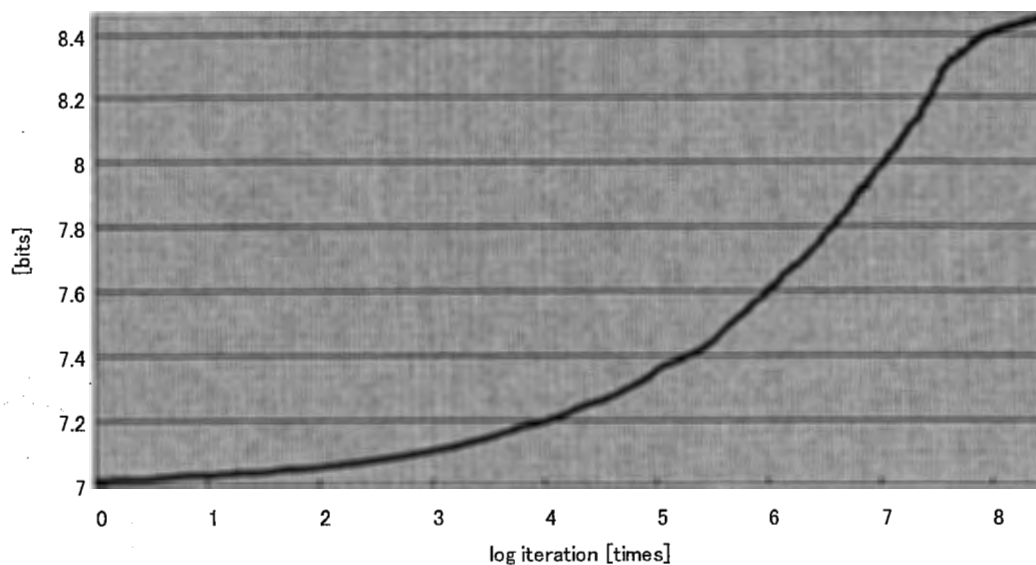


図:2 最適化過程におけるエントロピー変化

3.2 Initial+Final での実験

ここで用意した音韻は、Initial(21種類)、Final(37種類)の合計58である。

母集団は22220文あり、仮の文セットを7文とした。

最初の仮の文セット7文は、

1: sh ou zh ong sh ang de z i2 iu che cai pan l i1
qing zh i3 dou hao

2: z uo tian sh ang u i1 hui fu bu fen zh i3 j ve
dou hao

3: bing neng uo zhu tai de shou dou hao

4: x ia u z ai ba diao hu xi1 qi1 hou dou hao

5: zh uang kuang hai s uan un ding dou hao

6: zh i3 sh i3 hai iao guan

7: c ai neng pan duan sh i3 fou iu hou i1 zh eng
j v hao

であり、

この仮の文セット7文中にある音韻と音韻の出現回数は、表:3 に示した。

この仮の文セットのエントロピーは、5.0861になった。

a	2	ai	7	an	2	ang	2	ao	7
b	3	c	2	ch	2	d	11	e	3
ei	0	en	1	eng	3	er	0	f	3
g	1	h	13	i1	5	i2	1	i3	5
ia	1	ian	2	iang	1	iao	2	le	0
in	0	ing	3	iong	0	iu	2	j	2
k	1	l	2	m	0	n	2	o	0
ong	1	ou	11	p	2	q	2	r	0
s	1	sh	6	t	4	u	6	ua	0
uai	0	uan	3	uang	2	ui	1	un	1
uo	2	v	1	van	0	ve	1	vn	0
x	2	z	3	zh	7			others	0

表:3 音韻と音韻の出現回数(仮文セット中)

そして、文セット抽出処理を行った。

最終的な文セット7文は

1: j i1 u i g o n g z u o r e n v a n f e i i a o u o m e n m e i
 r e n m a i s a n v a n q i a n d e c h a z u o p i a o
 2: x i a z h o u j i a n g x i a n g b o c a i s i2
 3: t o n g n i a n l i u v e b i1 l i1 s h i3 j i a r u j v n t u a n
 4: t a l i1 l o n g x i u j i a f u g u o u a i s a n x i n
 5: u a i b i a o k e n e n g u i c a i h u a z h u a n g d u n h a o
 6: v s h i3 e r v e c h u s h a n g i n g d e i u m a o h a o
 7: v i e z i2 v a n p i n g h e n g u a n z h e n g d o u h a o

となった。

文セット中の音韻と音韻の出現回数は、表:4 に示す。

文セットのエントロピーは5.7229となった。

表:3 と 表:4 とを見比べてもわかるように、仮文セット中には出現回数が0という音韻が12個あった。しかし、処理後の文セット中にはすべての音韻がみられ、すべての音韻の出現回数も近い値となった。

a	2	ai	3	an	2	ang	1	ao	4
b	3	c	2	ch	2	d	4	e	3
ei	2	en	3	eng	3	er	1	f	2
g	2	h	5	i1	4	i2	2	i3	2
ia	3	ian	2	iang	2	iao	3	ie	1
in	1	ing	2	iong	1	iu	3	j	5
k	1	l	3	m	4	n	2	o	1
ong	2	ou	2	p	2	q	1	r	3
s	3	sh	3	t	3	u	3	ua	1
uai	2	uan	2	uang	1	ui	2	un	1
uo	4	v	2	van	3	ve	2	vn	1
x	4	z	3	zh	3			others	0

表:4 音韻と音韻の出現回数(文セット中)

	仮文セット	文セット
音韻環境種類数	46	58
総音韻環境数	135	139
エントロピー [bits]	5.09	5.72

表:5 仮文セット、文セットの品質(Initial+Final)

3.3 di-phone での実験

ここで用意した音韻は、C(37種類)、CV(97種類)の合計134種類である。(C:consonant, V:vowel)

例えば zhang には、zh+a と ang の音韻がある。そして、zh+a は CV に、ang は C に属する。

母集団は22220文あり、仮の文セットを25文とし文セット抽出処理を行った。

最初の仮の文セットのエントロピーは5.9753であった。

最終的な結果

文セットを25文は、

- 1: er b+e ei n+i1 iu s+o ong f+a a b+a an j+v v h+a ao
- 2: i1 s+e e l+i1 ie uai zh+a ang p+e ei s+i2 i2 sh+u uo
- 3: q+v v n+u uan
- 4: b+u u g+u uo l+i1 iu x+v ve q+i1 iang z+u uo t+i1 ian zh+i3 i3 z+a ai
k+a an sh+o ou n+e ei
- 5: s+u ui r+a an k+a ai f+a ang g+o ong ch+e e m+i1 in ing
- 6: q+v van l+o ong q+v van d+u ui ian j+i1 ian s+a an ang d+u ui
- 7: t+i1 ie l+u u j+v v m+u u q+i1 ian ing vn zh+u uang k+u uang
- 8: l+v v iu sh+o ou r+u u t+u u p+o o s+i2 i2 i1 van uai h+u ui r+e en
m+i1 in b+i1 i1
- 9: r+o ou xiar
- 10: r+u uo f+e ei g+o ong k+e e t+e e j+i1 ia d+o ou h+a ao
- 11: ch+i3 i3 d+a ao ia z+o ou l+i1 iao h+o ong x+i1 iong d+e eng s+a an
r+e en d+o ou h+a ao
- 12: z+a ai n+a a+er a-er

- 13: q+v van sh+e eng r+e en k+o ou ch+u u sh+e eng l+v v h+e e z+i2 i2
r+a an z+e eng l+v v
- 14: d+i1 i1 er c+u un
- 15: p+o o ua z+i2 i2
- 16: z+u uo z+ei ei c+o ong t+o ou c+a ai sh+i3 i3
- 17: sh+i3 i3 f+o ou c+e eng ve d+i1 ing
- 18: z+a ai n+a a+er a-er
- 19: g+u u u+i1 iahg m+e en c+o ong x+i1 iao j+i1 iu g+e en zh+e e m+a a
j+i1 iao h+u ua x+i1 iu h+u ua
- 20: x+i1 ing zh+i3 i3 b+o o b+o o d+e e f+u u r+i3 i3 b+e en
- 21: m+o o ter
- 22: t+a a m+e en d+e e j+i1 iao l+i1 ian n+i1 i1 er x+v vn d+eu ui b+u u
p+a a
- 23: ch+u u l+e e iong h+a ao x+i1 iang zh+e en q+i1 i1 ie d+a ai k+u uan c+u
un l+i1 iang h+e e zh+e eng ch+a ang x+i1 in z+e eng d+a ai k+u uan uai
- 24: c+i2 i2 c+i2 i2 zh+o ong h+u ua n+a an d+u un h+a ao n+v v l+a an d+a ai
b+i1 iao d+u ui
- 25: zh+u uan j+i1 ia p+i1 ing g+a ai ch+a an p+i1 in sh+a a ch+o ong g+u uang
p+u u

となった。

文セットのエントロピーは6.7256になった。

これもまた、monosyllable、Initial+Final と同様にすべての音韻の出現回数が近い値になり、音韻バランスがよくなった。唯一、音韻 u+o の出現回数だけが0であった。

	仮文セット	文セット
音韻環境種類数	98	133
総音韻環境数	552	394
エントロピー [bits]	5.97	6.73

表:6 仮文セット、文セットの品質(di-phone)

4. 考察

文音声データベースでは、できるだけ少ない数の文で、実際の中国語の発音に含まれる音声現象をすべて含んでいる事が望ましい。しかし、今回行った実験のなかで、monosyllable と di-phone での文セット抽出後、出現回数が 0 という音韻が、monosyllable の場合は 20 個、di-phone の場合は 1 個あった。この原因は、2.4 で説明したアルゴリズムに問題がある。このアルゴリズムはエントロピーを評価尺度としているため、すべての音韻が含まれているということを考えず、ただエントロピーが高くなる数文を抽出しているだけである。では、すべての音韻が含まれるようにするにはどうすればよいのかというと、まず1つに、このアルゴリズムをすべての音韻が含まれるということを前提条件におき、その中でエントロピーが高くなる数文を抽出してくるというアルゴリズムに変えるという方法である。もう1つは、文セット中に含まれていない音韻をリストアップして、その音韻を含むような文を付け加えるという方法である。どちらかというとな後の方が簡単に行えるのではないと思われる。

5. 最後に

今回抽出した文セットは、適当に仮文セットの文数を決めて行った結果であり、これが一番よいというわけではなく、文の数を減らせばもっとよい結果が得られるかもしれない。monosyllable に関して言えば、200文というのは少し多く、60から80文ぐらいにすればよかった。200文の時に出てこない音韻は20、80文にした時は約40である。よって、仮文セットを80文にして、出てこない約20の音韻を含む文を付け加えたほうが文の数を減らすことができ、200文の時よりもよくなる。