

満洲語古典語における母音調和の計算言語学的考察

坂上 温紀 坂井 優介 上垣外 英剛 渡辺 太郎

奈良先端科学技術大学院大学

sakajo.haruki.sd9@naist.ac.jp {sakai.yusuke.sr9,kamigaito.h,taro}@is.naist.jp

概要

満洲語における母音調和は他の言語における母音調和と異なる性質を持っており、今日まで様々な記述がなされてきた。17世紀ごろの満洲語に限れば、文字資料のみを用いてこの音声現象に向き合わねばならず、これには当時の文書を包括的に扱う必要がある。そこで本研究では Brown Clustering による母音・音節の分類と隠れマルコフモデルによる音韻環境と確率の計算を用いて母音のグルーピングと、母音 \bar{u} についての考察を行う。結果として、 $[\pm RTR]$ の調和があるが、母音の階層性によって調和が阻害されることがわかり、 \bar{u} は非円唇母音、あるいは弱い円唇性を持つ母音であると示唆された。

1 序説

満洲語古典語¹⁾は17世紀ごろの大清帝国で用いられていた、満洲・トゥングース語族に属する言語である。大清帝国で用いられていたこともあり、満洲語で書かれた文書は数多くあり[1]、東アジアのみならず、ユーラシア全体の歴史資料としても価値のあるものである。現在においても新疆ウイグル自治区や中国黒龍江省の一部地域で話されている。新疆ウイグル自治区チャブチャル・シベ自治県で話されているシベ語(満洲語シベ方言)の話者は3万人ほどとされており、これは満洲・トゥングース語族の中でも際立って多くの話者数である。このことは満洲語が満洲・トゥングース語族に関する言語学的研究にとっても重要な言語であることを意味する。

言語としての側面に注目すると、他のトゥングース語族と同様に、満洲語も母音調和がある。満洲語においては17世紀ごろの段階で母音調和が崩れているとも報告されている[2, 3]。この現象についてはたびたび議論されてきたが、その全容は未だに明らかにされていない。この現象に関する研究の多くは、いくつかの用例の収集とその分析に基づくもの

1) 以下、原則として満洲語と呼ぶ。

であるが、数多の用例に対する分析には自ずから限界が見えてくる。また、満洲語の母音は文字としては6種類あるが、音素としては5つであると論じられることもある[4]。このような論が生じるのは文字資料のみを用いて音声・音韻的な特徴を推測しなければならないためである。満洲語のような言語を研究するときには、利用可能な文字資料を余すことなく用いることで洞察を深めることが求められる。

そこで、本研究では母音のクラスターに基づく母音調和という現象を、計算機を用いて実際に母音をクラスタリングすることから改めて見つめ直す。これに、隠れマルコフモデルを用いて母音の共起確率の調査を加えた包括的な分析によって、満洲語の母音調和の根底にある法則を議論する。実験の結果、北方アジアの諸言語と同じように、満洲語も $[\pm RTR]$ による母音調和を持つと示された。さらに母音調和はaの後では崩壊することが示された。円唇性の調和の存在も確認され、これは母音 \bar{u} の円唇性に関する予測を導く。

本研究は満洲語の母音調和の実態を計算機科学的手法で迫るものであり、これにより意味ある解釈を与える事ができると示した。この点において、本研究は満洲語の計算言語学的研究の先駆けになるものである。

2 問題の所在

2.1 母音調和

音声学でいう音変化の中には「同化」という現象がある。同化とは、ある音が先行、または後続する音に近づくように変化をする現象である²⁾[5]。

母音調和はこの同化の一種で、語、あるいは形態素の中に現れる母音に制約がかかる現象である[6]。母音調和は満洲・トゥングース語族の諸言語やトル

2) 例えば、日本語には「読む」という動詞と「た」という接辞があり、「読む」+「た」は「読んだ」ではなく「読んだ」となる。これは「た」が先行する「む」という有声音に影響されて有声化した、同化の例である。

コ語などのいわゆるアルタイ型言語やフィンランド語などのウラル語族などで見られる現象である。

2.2 満洲語

本稿で扱う満洲語は17世紀から18世紀ごろのものであり、いわゆる有圏点満文である³⁾。満洲語はふつう満洲文字を用いて表記されるが、本稿では便宜上、いわゆる Möllendorff 方式によるローマ字転写を用いる⁴⁾。

本稿の研究対象となる満洲語の母音は文字としては6種類 (a, i, u, e, o, ü⁵⁾) あるが、音素としては6種類と考えられることもあれば、5種類と考えられることもある (付録 A を参照)。

一例として、6母音体系を採用するときの満洲語の母音の素性は次の通りである [7]。

- /i/: [-low, +coronal]
- /u/: [-low, -coronal, -RTR]
- /ü/: [-low, -coronal, +RTR]
- /e/: [+low, -RTR]
- /a/: [+low, +RTR, -labial]
- /o/: [+low, +RTR, +labial]

本稿では特に断りのない限りこの素性に基づいて考察を行う。

2.3 満洲語における母音調和

満洲語にも母音調和が存在しているとされるが⁶⁾、その規則には様々な議論がある [8, 9, 2]。陽性 (男性) 母音を /a/, /o/, /ü/、陰性 (女性) 母音を /e/、中性母音を /i/, /u/ とすることもあれば [10]、陰性 (女性) 母音に /u/ を加えることもある [11]。これは [±high] の調和 [4]、あるいは [±RTR] の調和とされる [7, 12]。陽性母音の内部では円唇性の調和が存在する [9, 12]。ただし、この分類に基づくだけでは満洲語の母音調和は説明できず、一定の傾向があったり何らかの制約が存在していたりする [10, 11, 13]。詳細は付録 B を参照のこと。本稿では計算機科学的手法により母音のグルーピングを行い、それを基に

3) 満洲語古典語、古典満洲語、満洲文語などと呼ばれる。
4) Möllendorff 方式は満洲文字が持つ情報をいくつか脱落させており、他の表記が用いられることもある [4, 2, 7, 3]。本稿では、用いるデータセットが Möllendorff 方式と極めて類似した方式を用いているためこれと同様の方式を用いる。
5) ü は u の長音のように見えるが、この転写は長音を表しているのではないことに注意されたい。u と表記されることもある。
6) ここで述べている満洲語の母音調和は母音調和ではなく、metaphony であるという主張もある [3]。本稿ではそれについて議論せず、一律に母音調和として扱う。

母音調和に関わる現象を説明する。

3 データセット

本稿で対象とするのは動詞における母音調和である。満洲語の母音調和は形態素内部でも、語幹に付加される接辞に対しても起こる (動詞に関する母音調和は付録 C を参照)。そこで、動詞に注目して分析することで、語全体としての母音調和と接辞の母音調和の両方を考察する。データセットの構築には、品詞情報の注釈付きデータセットである、ManPOS [14] の中から動詞と動詞接辞のペアを抽出した。

4 Brown Clustering による分析

満洲語の母音・音節にどのような関係があるのかをいわゆる Brown Clustering [15] を用いて分析する。このクラスタリングアルゴリズムは入力テキスト中の語それぞれが独立のクラスターに属するという状態から始まり、クラスターの併合を経て目的のクラスター数を持つ二分木を求めるものである。クラスターの併合は、その併合による相互情報量の損失を最小化するように行われる。ここで、 $n(w)$ 、 $n(w, w')$ をそれぞれ入力テキストにおける語 w 、bigram (w, w') の出現回数とし、 $n(c)$ 、 $n(c, c')$ をそれぞれ入力テキストにおけるクラスター c 、bigram クラスター (c, c') の出現回数とすると、クラスタリングの品質は以下で定式化される。

$$Q(C) = \sum_{c, c'} \frac{n(c, c')}{n} \log \frac{n(c, c')n}{n(c)n(c')} + \sum_{w'} \frac{n(w')}{n} \log \frac{n(w')}{n} \quad (1)$$

本稿では式 1 を用いつつ、最適化の手法に改良を加えたアルゴリズム [16] を用いる。

3種類のテキストに対してクラスタリングを行う。taci-mbi という動詞語幹・接辞を用いた入力は次の通り。

1. 母音のみ (例: aii)
2. 動詞語幹・動詞接辞それぞれの音節⁷⁾ (例: ta ci mbi)
3. 動詞語幹の母音と動詞接辞 (例: ai mbi)

入力 1 によって母音の分類を観察し母音調和の基本原則を考察する。入力 2 によって子音との組み合わせ

7) 語幹と接辞に分けており、さらに文字に基づいて分割しているため、実際の音節とは異なることに注意されたい。tacimbi は本来 ta-cim-bi と区切られる。他にも tiyan という文字列は CVVC という音節構造の /tian/ と表される [17] ように、表記と音が一致しないことがある。

表 1: 動詞に含まれる母音のみを対象とした分類の結果

(a) クラスタ数: 6		(b) クラスタ数: 4	
path	vowel	path	vowel
00	e	0	e, u
01	u	10	i
100	i	110	o
101	o	111	a, ū
110	ū		
111	a		

表 2: 音節の分類の結果 (クラスタ数: 20)

path	vowel
000	he, ke
0010	ci, re, ho, ro, nu
0011	fi, me, mbi, ki, mbihe
010	ra, šo
011	ha
100	ji, ne, te, du, le, ce, de, je, ma
1010	la, fa, ca, ri, ša, bo, b
10110	da, si, ka, sa, na, so, kūm
10111	li, sim, hi, lo, cu, lu, ko, mi
110	ge, u, ga, hen, ban, sin, bu, fu, be
1110	a, i, ja, jo, go, fak, šang, dai
111100	wa, ya, o, gai, su, to, heng, ger
111101	bi, gi, tu, ni, e, we, do, se
111110	kū
1111110	ung, gū, gu, ju, ok, bar, nong
1111111	ba, ta, hū, bai, hūl, las, us

せによる母音の分類を観察する。入力 3 によって、語幹の母音連続の分類を観察する。入力 1 はクラスタ数を 6 または 4 とし、入力 2、3 はクラスタ数を 20 とした。この 3 種類の入力を用いた結果は順に表 1、表 2、表 3 である。

表 1a を見ると、6 つの母音は大きく (1) e, u (2) i, o, ū, a の 2 グループに分類される。表 1b も表 1a と似た分類を示しているが、i がより根に近いノードにある点が異なる。表 2 からは明瞭な分布を得ることは難しいが、その傾向を見て取れる。表 3 は 00 から始まるクラスタが (1)、01 から始まるクラスタが (2) を表している。

5 HMM によるモデリング

隠れマルコフモデルを用いてある音韻環境 e が与えられたときのある母音 v への遷移確率を求める。この遷移確率を基にして、母音のグルーピングと母音調和の法則を議論する。隠れマルコフモデルの状態数は 10 とする。入力は第 4 節で用いた入力 1 を用いる。6 種の母音の bigram、trigram の対数確率は表 4 の通りである。

表 3: 動詞語幹の母音と動詞接辞を対象とした分類の結果 (クラスタ数: 20)

path	vowel
000	i
0010	ue
00110	ee, ui, e, ei, eie, iue, u, ū
00111	eu, eue, ie, eee, uue, uee, eoe, iee
010	oo, ua, ūi, aua, ūa, aia, iia, iu, ooa, ūaa
01100	ai, oi, aū, au, aai, iii, uu
01101	a, ii
01110	o
01111	aia, aaa, uaa, uaia, aiaa, uia
10	he, hebi, hekū
1100	ra
1101	ha, kakū, habi
11100	me, ka, rahū
11101	ho, ro, hobi, ko, hoi, tolo
11110	re, ke, tala
111110	fi, ki, mbihe
1111110	rakū
11111110	mbi
11111111	ci, kimi, cibe

6 考察

6.1 満洲語の母音調和

Brown Clustering の結果から、母音は (e, u) と (i, o, ū, a) に分けられる。i を除けば、(e, u)、(o, ū, a) という分類は [±RTR] で区別される。これは満洲・トゥングース語族によく見られる対立であり、満洲語の母音調和もこれに基づくものであることが示される。階層的に見ると、まず [±RTR] で大きく分けられている。この内、[-RTR] のグループは [±labial] で分けられる。[+RTR] のグループは表 1b に従えば、(i)、(o)、(a, ū) と分けられる。i については、表 2 でも表 3 でも、どのクラスタにも見られることから中性母音と考えてよいだろう。動詞接辞における母音調和は a/o/e の対立が主であるから、残る 2 つのグループは a/o の区別である。表 4 を見ると、o は先行する母音が o のみのときには起こりやすく、a/o は [+RTR] における円唇性の調和によって決まると考えられる。

また、表 4 から [-RTR] の調和は先行する母音に a が含まれないときに限り生じることもある⁸⁾。a が含まれていないときにも、uū という母音の組み合わせに対しては a の確率が最も高い。このような

8) [±RTR] の母音調和という仮定の下で、[-RTR] の母音と [+RTR] の母音の共起は母音調和が崩れていることを意味する。

表 4: 動詞に含まれる母音のみを対象とした対数確率 (隠れ状態数: 10)。たとえば最も aa 行 e 列の数字 -3.38 は aae の対数確率を表す。

	a	ū	o	i	u	e
a	-1.53	-4.45	-23.22	<u>-1.82</u>	-4.54	-4.9
aa	-2.19	-6.31	-16.45	<u>-2.7</u>	-7.44	-3.38
aū	-5.11	-9.27	-19.39	<u>-5.61</u>	-10.37	-6.32
ao	-26.29	-29.64	-25.41	<u>-24.36</u>	-25.88	-24.03
ai	-2.47	-6.61	-16.74	<u>-2.99</u>	-7.73	-3.67
au	-5.2	-9.36	-19.46	<u>-5.71</u>	-10.45	-6.39
ae	-5.56	-9.38	-19.83	<u>-6.07</u>	-10.83	-6.74
ū	-4.83	-7.75	-15.5	<u>-5.05</u>	-7.51	-6.76
ūa	-5.48	-9.61	-19.75	<u>-6.0</u>	-10.74	-6.68
ūū	-8.41	-12.57	-21.62	<u>-8.92</u>	-13.49	-9.59
ūo	-19.9	-21.91	-19.95	<u>-16.8</u>	-17.8	-16.02
ūi	-5.76	-9.86	-18.12	<u>-6.22</u>	-10.55	-6.74
ūu	-8.49	-12.58	-17.8	<u>-8.71</u>	-10.98	-8.77
ūe	-8.57	-11.76	-17.19	<u>-7.93</u>	-10.41	-7.47
o	-4.41	-8.9	<u>-3.26</u>	-2.85	-9.0	-6.85
oa	-5.2	-7.53	-17.18	<u>-5.47</u>	-8.74	-6.38
oū	-9.61	-11.98	-23.48	<u>-10.09</u>	-14.88	-10.77
oo	-4.9	-9.06	<u>-4.59</u>	-4.68	-9.56	-4.49
oi	-3.78	-7.22	-9.61	<u>-3.99</u>	-8.27	-4.17
ou	-9.76	-13.89	-20.29	<u>-10.18</u>	-13.34	-10.59
oe	-8.54	-9.07	-14.99	<u>-8.06</u>	-12.81	-7.76
i	-3.46	-6.35	-7.87	<u>-3.19</u>	-4.83	-3.18
ia	-4.15	-6.82	-18.41	<u>-4.65</u>	-9.44	-5.33
iū	-7.08	-10.56	-18.24	<u>-7.48</u>	-11.16	-8.1
io	-9.53	-13.68	<u>-9.22</u>	-9.28	-13.41	-9.07
ii	<u>-4.35</u>	-7.18	-13.67	-4.36	-7.62	-4.25
iu	-7.18	-10.73	-14.04	<u>-6.1</u>	-7.27	-5.46
ie	-5.82	-7.6	-13.42	<u>-4.36</u>	-6.66	-3.74
u	-3.57	-6.41	-11.3	<u>-3.02</u>	-4.42	-2.76
ua	-4.22	-7.94	-18.49	<u>-4.74</u>	-9.49	-5.42
uū	-7.16	-11.16	-17.78	<u>-7.59</u>	-10.82	-8.02
uo	-16.83	-17.75	-20.31	<u>-12.59</u>	-13.56	-11.79
ui	-4.37	-7.81	-14.03	<u>-4.19</u>	-7.2	-3.92
uu	-7.18	-10.48	-13.56	<u>-5.7</u>	-6.79	-5.0
ue	-5.54	-7.86	-12.94	<u>-3.94</u>	-6.18	-3.29
e	-6.94	-7.96	-7.35	<u>-2.95</u>	-3.93	-2.17
ea	-8.02	<u>-8.26</u>	-21.16	<u>-8.26</u>	-12.82	-9.02
eū	-10.54	-12.3	-17.19	<u>-9.09</u>	-10.43	-8.64
eo	-9.03	-13.17	<u>-8.71</u>	-8.76	-12.55	-8.54
ei	-5.63	-7.7	-12.87	<u>-4.11</u>	-6.66	-3.49
eu	-9.44	-10.38	-12.94	<u>-5.22</u>	-6.19	-4.42
ee	-5.06	-7.26	-12.33	<u>-3.34</u>	-5.57	-2.68

制約から a > e > ū > o, u という階層構造が推察される。この階層構造の上位の母音と [+RTR] で調和し、[+RTR] についてはさらに円唇性の調和が起こることがある、と言える。

6.2 第 6 の母音 ū について

第 2.2 節で述べたように、ū は u と同一の音素であるという考えがある。しかし、第 6.1 節で論じたよ

表 5: 音節とその使用例 (抜粋)

音節	使用例
gū	gūnici, gūnifi, gūnirakū, gūwaliyaka
ung	unggihe, unggifi, unggire, ungnembi
gu	gurime, gurire, guwembi, guculefi

うに、ū と u は同一クラスには属していない。さらに、表 4 は ū と u が異なる振る舞いをしていると示している。これらの結果から、この 2 つの母音が異なる音素であると言える。一方で、表 2 では gū は u を音節核にもつ音節と同じクラスター (1111110) に属している。gū という音節の使用例 (表 5) を見ると、“gūni-” という動詞での使用が大部分であった。同じクラスターに属する音節 (ung, gu) を見ると、これらの音節の後には i を音節核にもつ音節が続いている。Brown Clustering は bigram 頻度を用いる手法であるから、後続する音節核を共有しているため、gū/ung/gu が同じクラスターに分類されたとわかる。そこで、より広く分布を見ると、ū を音節核に持つ音節が属するクラスター (111) は a が支配的であり、ū は a と同じグループに属することを示している。

第 6.1 節で [+RTR] には円唇性の調和があると論じたが、ū のみ、または ū と o のみを持つ環境において、o が現れる確率は低い (表 4)。[+RTR] の円唇性の調和は [-low] にのみ起こる [7] ことを踏まえると、ū が円唇性を持たない、あるいは円唇性が弱い可能性が考えられる。

以上のことから、ū は u とは異なる音素の母音であり、非円唇母音、あるいは o より弱い円唇性を持つ母音と推察される。

7 結論

本稿では動詞語幹と動詞接辞を用いて、Brown Clustering と隠れマルコフモデルによって満洲語の母音調和の原理の解明を試みた。その結果、[±RTR] の調和があることがわかり、さらに調和 (あるいは単に同化) を引き起こす母音の階層構造の存在が示唆された。また ū が非円唇母音か弱い円唇性を持つ母音である可能性を示した。本研究は満洲語の計算言語学的研究の先駆けとなるものであり、他の品詞への対象の拡大や通時的な分析への発展に繋がるものである。本研究を通じて、文字資料しかない満洲語の研究に対する一つのアプローチとして、計算言語学的手法を用いた満洲語研究が進むことを望む。

参考文献

- [1] 长山. 《满文原档》语言文字价值研究. 故宫博物院院刊, 2023.
- [2] 早田輝洋. 満洲語の多様な形態素-ngge の文法的位置付け. 言語研究, Vol. 148, pp. 33–60, 2015.
- [3] 王海波. 満洲語現代方言における母音調和. 北方言語研究, Vol. 10, pp. 135–156, 03 2020.
- [4] 早田輝洋. 満州語の母音体系. 九州大学言語学論集, Vol. 2, pp. 1–10, 2003.
- [5] 入江浩司. 同化・異化. 斎藤純男, 田口善久, 西村義樹 (編), 明解言語学辞典. 三省堂, 2015.
- [6] 斎藤純男. 母音調和. 斎藤純男, 田口善久, 西村義樹 (編), 明解言語学辞典. 三省堂, 2015.
- [7] Seong Ko. **Tongue Root Harmony And Vowel Contrast In Northeast Asian Languages**. PhD thesis, Cornell University, August 2012.
- [8] Josh Ard. Vowel harmony in Manchu: A critical overview. **Journal of Linguistics**, Vol. 20, No. 1, pp. 57–80, March 1984.
- [9] Xi Zhang and B. Elan Dresher. Labial Harmony in Written Manchu. **Saksaha: A Journal of Manchu Studies**, Vol. 1, , 1996.
- [10] 河内良弘, 清瀬義三郎則府. 満洲語文語入門. 京都大学学術出版会, 2002.
- [11] 津曲敏郎. 満洲語入門 20 講. 大学書林, 2002.
- [12] Andrew Jonathan Joseph. **The Historical Phonology of Manchu Dialects**. PhD thesis, Cornell University, Ann Arbor, United States, 2018.
- [13] 爱新 觉罗 乌拉熙春. 满语语法. 内蒙古人民出版社, 1983.
- [14] Sangah Lee, Sungjoo Byun, Jean Seo, and Minha Kang. ManNER & ManPOS: Pioneering NLP for endangered Manchu language. In Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue, editors, **Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)**, pp. 11030–11039, Torino, Italia, May 2024. ELRA and ICCL.
- [15] Peter F. Brown, Vincent J. Della Pietra, Peter V. deSouza, Jenifer C. Lai, and Robert L. Mercer. Class-based n -gram models of natural language. **Computational Linguistics**, Vol. 18, No. 4, pp. 467–480, 1992.
- [16] Percy Liang. Semi-supervised learning for natural language. Master's thesis, Massachusetts Institute of Technology, 2005.
- [17] 舞格. 滿漢字清文啓蒙, 1732.

A 5 母音体系について

5 種類と考えるときには、/ū/ という音素は存在せず、u と ū は同一の音素の異音であるとされる [4, 2]。この根拠として、漢字音の表記や文字の形、母音調和を挙げている。

B 満洲語の母音調和について

以下に満洲語の母音調和に関する説明を挙げる。グルーピングとしての違いは ū や u の取り扱いが異なる場合がある。グルーピングの説明としては [±ATR] か [±RTR] の調和と円唇性の調和について述べられることが多い。

Zhang & Drescher (1996) [9] 以下のような母音目録の下で、舌の位置に関する調和の次に円唇性による調和を考える。円唇性の調和はハルハモンゴル語やエヴェンキ語にも見られる。

- /i/: [+ATR, +labial]
- /u/: [+ATR, -labial]
- /ū/: [-ATR, -labial]
- /e/: [+ATR, +labial]
- /a/: [-ATR, +labial]
- /o/: [-ATR, -labial]

津曲 (2002) [11] 以下の3つのグループに基づく説明される。(i, u) は中立 (中性母音) としている。

1. a, o, (k, g, h のあとの) ū
2. e
3. i, u

(a, o) の選択に関しては o を含む音節のあとには o が続きやすい、という傾向があると述べている。

河内 & 清瀬 (2002) [10] 以下の3つのグループに基づく説明される。(i, u) は中立 (中性母音) としている。

1. a, o, ū
2. i, u
3. e, (k, g, h のあとの) u

早田 (2003) [4] 動詞連体形接辞の母音調和を

- 直前の母音が a であれば a
- 語幹の母音が [+high] (あるいは [+ATR]) であれば e

と記述している。

Ko (2012) [7] 以下の素性の下で、[±RTR] の調

和と [+low] 内の円唇性の調和について述べている。

- /i/: [-low, +coronal]
- /u/: [-low, -coronal, -RTR]
- /ū/: [-low, -coronal, +RTR]
- /e/: [+low, -RTR]
- /a/: [+low, +RTR, -labial]
- /o/: [+low, +RTR, +labial]

C 満洲語の動詞

満洲語の動詞は動詞語幹に動詞語幹を付加することにより作られる。たとえば taci- 「学ぶ」という動詞語幹には -mbi/-ha/-re などの動詞語幹が付加される。動詞語幹の中には -ha/-ho/-he のような母音交替を起こすものがあり、動詞語幹が持つ母音によって母音が交替する。