

## 研究ノート

## 比較言語学における数学的方法

吉 田 知 行

## 目次

- 1 はじめに
- 1.1 比較言語学の始まり
- 1.2 日本語の起源問題
- 1.3 新しい数理的方法
- 2 比較言語学
- 2.1 比較言語学
- 2.2 日本語の起源問題
- 2.3 日本語の謎と不思議
- 3 比較言語学における数理的方法
- 3.1 ボリアの二項検定法
- 3.2 オズワルトのシフト法
- 3.3 完全シフト法
- 3.4 上古日本語・中期朝鮮語・アイヌ語の比較
- 3.5 言語年代学
- 4 あとがき

## [要旨]

比較言語学に役に立つと思われる数学的な手法を解説する。第1章では、比較言語学の始まりから、近年登場してきた数理的方法の概要を紹介する。また、日本語の起源探求の困難さと、日本人の起源をめぐる最近の考古学とDNA 人類学から見えてきた日本人の起源の概要を紹介する。第2章では、日本語の起源について比較言語学からの研究を紹介する。第3章では、比較言語学における数理的方法、とくに一致数で言語間の近さを測るボリアの二項検定法とオズワルトのシフト検定法を紹介する。また言語年代学の基本公式の改良をする。例として、日本語・アイヌ語朝鮮語を比較する。さらなる計算例として、これらの言語とインドネシア語・カンボジア語などとの比較をする。

## 1 はじめに

## 1.1 比較言語学の始まり

比較言語学は、18世紀末の東洋学者ジョーンズ卿 (1746-1794) の研究に始まる。

William Jones “The Sanscrit Language”(1786)

彼は、今から2000年以上前から伝わってきたインドのサンスクリット語 (梵語) の単語の中に、ギリシア語やラテン語に似たものを数多く見出した (表1)。この観察から、彼は「3つの言語は、共通の源 (インドヨーロッパ祖語) から発した」と考えるようになった。

表1 英語・サンスクリット語・ギリシア語・ラテン語の単語

言語	二	三	十	父	母	兄弟
eng	two	three	ten	father	mother	brother
san	dvi	tri	dasa	pitar	matar	bhratar
gre	duo	treis	deka	pater	meter	phrater
lat	duo	tres	decem	pater	mater	frater

その後多くの言語学者の努力によって、ヨーロッパ、イラン、インドのかんりの言語、さらにヒッタイト語やトカラ語までを含むインド・ヨーロッパ語族 (印欧語族ともいう) が確立した。比較言語学では、言語の語族への分類、ふたつの言語が同系かどうか、言語がどのように分岐し変化してきたかが研究され、さらに祖語の復元までが試みられてきた。

キーワード：一致率検定, 二項検定, 比較言語学, 言語年代学, DNA 人類学, 日本人の起源, 日本語の起源

現在では、分子人類学による遺伝子と語族の対応の研究が盛んである。

## 1.2 日本語の起源問題

ひるがえって、日本語の起源問題については、100年以上の研究の歴史がありながら、日本語と同系と確定した言語は存在しない。日本語の系統は不明で、孤立した言語というのが大方の言語学者の考えのようだ(服部四郎 1959)。日本語以外にアイヌ語と朝鮮語、それにギリヤーク語も同系の言語を持たない孤立した言語である。何故、極東にそのような言語が集中しているかも不思議である。

日本語の研究を困難にする理由がいくつかある。

1. 同系をうかがわせる言語がない。
2. 日本語は他の言語と分岐してから相当長い時間が経過していて、比較言語学の適用範囲(6~7千年前)を超えていると考えられる。
3. 孤立した言語には比較言語学が使えない。
4. 極東には古い言語資料が乏しい。
5. 日本語の成立過程は相当複雑である。縄文時代の末期(約2500年前)に、大陸から弥生人の言語が流入し、2000年ほど前から中国語や朝鮮語が流入し、近年ではヨーロッパの言語が日本語の中に大量に混入している。
6. 日本語のクレオール語や流入説もある。この場合も、従来の比較言語学の方法が使えない。
7. 日本語の起源問題はトンデモの世界で、まともな言語学者は参入しづらい。万葉集が韓国語で読めたとか、日本の起源はインドの言葉だとか、言語学の議論とは思えない。

本論文では触れないが、言語類型論の立場

からの研究もある(松本克己 2007)。例えば、日本語とその周辺にある言語の語順の分布から、日本語の祖語もSOV型であったろう。

日本人と日本語の起源問題研究は、片方が進展すれば他方も進展する。

海外では、特定の文化、とくに言語と遺伝子を対応付けた研究が盛んである。とくに、Gray-Atkinsonによるインドヨーロッパ語族の系統樹の作成(2003)や、Grayによるオーストロネシア系言語の系統樹の作成(2009)がよく知られている。これも言語と遺伝子を結びつける研究の応用である。

関心のある人類学者はいるが、日本ではまだそのような研究が少ない。そんな中で、崎谷満の研究が目立つ。一連の論文著作で、同氏は極東の文化、言語、遺伝子を結びつけた研究を行っている。とくに日本語の方言の起源についての研究が興味深い。この方面の研究は今後大いに期待出来るテーマである。

この数年の遺跡発掘と年代測定、さらに分子生物学の急速な発展により、日本人の起源についてかなりのことが分かってきた。ホモサピエンスの出アフリカから日本列島到達までの経過を大ざっぱにまとめると次のようになる。ただし年代など見解の一致していない部分もある。

1. 約7万年前。ホモサピエンスの集団がアフリカを出た。
2. 約4万年前。ホモサピエンスが東アジアに到達した。南ルート説、北ルート説がある。
3. 3万8000年前。日本列島の後期旧石器時代の始まり。大陸から日本列島への渡来ルートは、サハリン・北海道ルート、朝鮮半島・対馬ルート、台湾・琉球ルートが考えられる。
4. 約1万5000年前。氷河期が終わり、土器や石器の製造技術を身につけた後期旧石器時代人が縄文文化を広げる。縄

文人は、各地で独自の文化を発展させる。

5. 約2千500年前。大陸から弥生人が渡来し、稲作など弥生文化を広げる。

つまり、日本人の起源は相当古く、3万5千年以前の後期旧石器時代までさかのぼることが分かってきた。結局のところ、日本語の起源も、比較言語学の適用範囲を大きく超えている。

### 1.3 新しい数理的方法

語彙の比較による比較言語学には、分岐してから6000～7000年程度の適用限界がある。また同系でない言語には使えない。しかし近年比較言語学のための新しい方法が開発されている。それは数学とくに確率・統計の考えを使うことである。本論文では、ポリア・安本の二項検定法とオズワルトのシフト検定法を紹介する。

本論文の元になったのは、吉田による次のふたつの原稿による：

- ・「数学の応用事例—比較言語学への応用」サイエンストピックス，北海道大学理学部ホームページ(2005)。『季刊邪馬台国』に転載。
- ・「分割表の一致率検定のための正確確率法と比較言語学への応用」統計数理研究所での講演（2007/02/22）。

計量比較言語学については、以下の書籍を参考にした。

安本美典・野崎昭弘『言語の数理』筑摩書房(1968)

## 2. 比較言語学

この節は、著者の古い講義ノートを元にした前出「サイエンストピックス」(2.3.10節)

をほぼそのまま収録した。文体は整えた。今となっては記述が古く、年代についてはいくつかの誤りがあるが、そのままにしてある。

### 2.1 比較言語学

ふたつの言語が同系かどうかを判定する基準として、文法、音韻、語彙の類似が考えられる。英語を学んですぐ気がつくのは、日本語と英語の語順の違いである。日本語はSOV型（ここでSは主語、Oは目的語、Vは述語）であるのに対し、現代英語や中国語はSVO型である。ではインド・ヨーロッパ語族に属する言語がすべてSVO型かという、そんなことはまったくなくて、肝心のラテン語はSOV型（古代英語や古代ギリシア語もどちらかというSOV型）だし、ウェールズ語やアイルランド語などのケルト系言語はVSO型である。現代ドイツ語はどちらかというSVO型だが、関係代名詞の中はSOV型である。世界の言語の中にはOSV型やVOS型の言語もある。

ちなみに、古代英語は、名詞には男性女性中性の区別があり、名詞や動詞の語形変化も激しいなどドイツ語に似た言語だった。英語は、科学の世界で世界共通語の役割を果たしつつあるが、世界の言語の中ではかなり変わった言語である。子音（24種類の音素）や母音（短母音7種類、そのほかに長母音、二重母音、三重母音）がやたらと多い。また綴り字と発音はかなりずれているし、さらに膨大な数の語彙（数十万?）がある（これは日本語も同じ）。

英語は、元のインドヨーロッパ語族に比べると、ある意味で退化した言語といえよう。名詞の男性・女性・中性の区別が完全になくなり、格変化もほとんど消滅した。複数形は若干の例外を除けば単に's'か'es'を語尾につけるだけで良くなったし、ドイツ語やロシア語のような動詞の屈折変化も乏しくなった。英語は、ゲルマン系言語の上に北欧の言

表2 ヨーロッパの言語の数詞 (文字表記)

	英語	スウェーデン語	デンマーク語	オランダ語	ドイツ語	フランス語	スペイン語	イタリア語	ポーランド語	ハンガリー語
1	one	en	en	een	ein	un	uno	uno	jeden	egy
2	two	tra	to	twee	zwei	deux	dos	due	dwa	ketto
3	three	tre	tre	drie	drei	trois	tres	tre	trzy	harom
4	four	frya	fire	vier	vier	quatre	cuatro	quattro	cztery	negy
5	five	fem	fem	vijf	funf	cinq	cinco	cinque	piec	ot
6	six	sex	seks	zes	sechs	six	seis	sei	szesc	hat
7	seven	sju	syv	zeven	sieben	sept	siete	sette	siedem	het
8	eight	atta	otte	acht	acht	huit	ocho	otto	osiem	nyolc
9	nine	nio	ni	negen	neun	neuf	nueve	ove	dziewiec	kilenc
10	ten	tio	ti	tien	zehn	dix	diez	dieci	dziesisc	tiz

語やラテン語、フランス語などがかぶさってきた。ごった煮のような状況では、名詞の性や格変化、動詞の複雑な屈折変化のような「無駄」は切り捨てられてしまったのだろう。かわりに前置詞と助動詞が発達し、語順(SVO型)に厳しい制限が付くようになった。このように、文法も変化するので、文法の類似だけでは同系の証明にならない。

比較言語学で同系の証明をどうするかというと、それは「音韻対応の法則」(または「音法則」)の成立によってなされる。例えば、英語とドイツ語では、drink↔trinken, think↔denken, day↔tagから直ちに見て取れるd↔tやt↔dのような規則正しい対応がある。ヨーロッパのいくつかの言語の数詞のリストをあげておく(ただし文字表記で補助記号は省略)。

一見して明らかなように、多くの言語で数詞はよく似ている。ハンガリー語をのぞくすべての言語で、6, 7はs音で始まり、3はtまたはdで始まっている。ポーランド語を除くすべての言語で1は母音で始まっている。また、フランス語よりドイツ語の方が英語に近い印象を受ける。ドイツ語と英語が分かれたのは今から2000年前、フランス語とドイツ語が分かれたのが5000年前といわれている。ただしハンガリー語だけは他の言語との類似性

が感じられない。実際、他の言語はすべてインド・ヨーロッパ語族なのに、ハンガリー語だけはウラル語族である。

## 2.2 日本語の起源問題

インドヨーロッパ語族やマライポリネシア語族などで、比較言語学は大きな成功を収めてきた。これに対して、日本語の系統関係は全く不明で、日本語との同系関係が証明された言語は沖縄語(沖縄方言)だけである。日本語は○○語と同系であるとか、古代日本語が△△語で読めるとかいう本が結構出ているが、まずは眉唾物と考えて間違いない。当然ながら、そのような説は、ほとんど(おそらく全部)が間違っているはずである。単語の範囲を無制限に広げ、個々の単語の意味も曖昧にするなら日本語は英語と同系であるとさえ言える。例えば、「斬る」と「kill」,「坊や」と「boy」,「そう」と「so」,「負う」と「owe」,「道路」と「road」,「字を引く書なり」と「dictionary」のように意味と発音の似た単語はたくさん挙げることができる(清水義範『序文』)。

このような語呂合わせを防ぐために、比較言語学では、言語の同系の判別に「音韻対応の法則」という厳しい条件を課している。また、比較言語学的方法で復元した単語の祖形

を比べて、日本語は××語に同系だ、という議論もよく見られる。専門家のやることなので軽々しく批判は出来ないが、何か一貫性のない名人芸に陥っているような感じを受けることがある。祖語復元のアルゴリズムがあるなら納得できるのだが。

日本語と他言語との同系関係の有無は、100年も前から比較言語学の専門家が研究し続けてきた。それでも定説といえる同系関係を見出だせないなら、もはやこの地球上に日本語（含沖縄方言）と同系の言語は存在しないとさえ考えられる。

なら、日本語は全く孤立した言語なのかというとなんかそうじゃない。そもそも人類（ホモサピエンス）の祖先がアフリカを出たのが15万年前、アジア人とヨーロッパ人が分かれたのが5万年前といわれている。日本語や新大陸の言語を含む東アジアの言語は、2万年ほど前に分岐したと考える言語学者もいる。そうすると日本語も2万年前にさかのぼれば、東アジア人の共通言語にたどり着くはずである。しかし残念なことに、分岐の時期がいつであれ、音韻対応の法則を見出す方法では（いや、おそらくいかなる方法でも）、これほど古く分岐した言語の同系を証明することは不可能に近い。音韻対応の法則を見出すという方法で、5000年以上前に分岐した言語の同系関係の証明や祖語の復元が可能である。しかしそれも、比較出来る多くの言語が周囲にあり、それらの古い記録が残っているという好条件に恵まれてのことである。現代の英語とネパール語だけを比べて、それらが同系であると証明するのは困難であろう。同じインドヨーロッパ語族に属しながら、ネパール語の語順などは英語より日本語に似ている。

日本語が他の言語から分かれたというのなら、その年代として、弥生時代の初め（2400～3000年前？）と縄文時代の初め（12000年以上前）と考えられる。その年代が弥生時代

の開始時期で、日本語の元になった言語の子孫がどこに残っていたら、言語学者が見つけないはずはない。反対に、日本語の祖語が1万年以上前に他の言語から分岐し、日本列島内で現代の日本語に変化していったというなら、もはや同系関係を証明するのは困難である。結局、日本語と同系な言語はもう残っていないか、残っていても分岐年代が古すぎるなどの理由で、比較言語学の方法では同系の証明ができない、と考えるのが妥当である。特に中国語以外のアジアの言語に、古い文字資料の少ないことが致命的である。

やや絶望的な状況だが、まだ探求の道はある。日本列島周辺に分布している言語と比べると、日本語は、中国語以外のアジアの諸言語の中ではきわめて標準的で平凡な音韻や文法体系を持っている。特に朝鮮語やモンゴル語の文章は、単語を対応する日本語の単語に置き換えるだけでほとんどそのまま日本語の文章になる。つまり日本語ときわめてよく似た文法体系を持つ北方系の言語がある。そのため、日本語も朝鮮語のアルタイ（トルコ語、モンゴル語など）説が有力である。日本語と朝鮮語が同系だとしても、その分岐年代は今から6700年ほど前、という言語年代学の計算結果もある。文法は北方的だが、音韻や語彙は、南方系だという言語学者も多い。伝達の経路がどうであれ、稲作は明らかに南方系の文化要素なので、稲作とともに南方系の言葉が伝わったと考えるのはきわめて自然である。

また、クレオール語（混成言語）説や流入説もある。日本列島にはいろいろな時代にいろいろな民族がいろいろな文化を持ち込み、そのたびに新しい言葉が流れ込んで来た。やまとことばを核として成立した日本語には、千年以上の間に膨大な量の漢語が加わり、現代でもカタカナ語として大量のヨーロッパの言語が流れ込んでいる。日本語が他言語の単語を特別受け入れやすい言語なのかもしれない



いが、有史以前からこのような状況だったとすれば、流入説の方に説得力がある。いずれにせよ、定説のなさ、日本語の起源問題の難しさと、この問題が従来の比較言語学の方法の射程外にある可能性を示唆する。問題解決には新たな方法が必要である。

## 2.3 日本語の謎と不思議

ここは吉田 (2005) の10節の内容をそのまま転載する。内容は、言語学的に厳密な話ではなく、個人的な感想がほとんどである。

日本語にはたくさんの不思議がある。例えば次のようなものである。

- (a) 数詞の語頭子音の倍数構成 (1:hi と 2:hu, 3:mi と 6:mu, 4:yo と 8:ya, 5:itu と 10:to)。確率計算から、この対応は偶然ではない。しかし、何千とある言語の数詞を調べても、倍数構成の痕跡を持つ言語はハンガリー語などきわめてわずかである。このような数詞の倍数構成は、日本列島の中で生じたのだろうか、それとも外で生じたのだろうか。そもそも日本語の和数詞はどこでどのようにして生まれたのだろうか。
- (b) 高句麗語の数詞。『三国史記』という高句麗・百濟・新羅三国の正史を表した12世紀高麗の書物の中に、高句麗の地名がたくさんある。広辞苑の編者として有名な新村出は、その中に数多くの日本語と思われる単語を見つけている (1913)。例えば、「三紛縣 (一云密波兮)」「五谷郡 (一云于次吞忽)」「七重縣 (一云難隱別)」「十谷縣 (一云徳頓忽)」とある。つまり数詞「三」「五」「七」「十」の一云 (読み) が「ミツ」「ウチャ」「ナン」「トク」だという。こうなると両言語での類似は明らかである。よくこのようなことを発見したと驚く。数字以外の文字についても、日本語との対応がいくつか発見

されている。これらの語が本当に高句麗語 (朝鮮語系言語) の単語なのか疑問が残るが、それならいったいどのようにして日本語の単語があのような場所にまで入り込んだのだろう。

- (c) 日本列島に到来した言語。日本列島には1万年以上前から様々な文化が流れ込んできた。例えば、最近の考古学の研究で3500年前の縄文時代にはすでに稲作がある程度行われていたことが分かっている。それなら、そのような縄文稲作をもたらした人々はどんな言葉を話していたのだろう。

追加コメント。(a) 数詞が倍数構成で作られている言語としては他に、カナダ西岸最南部のハイダ語がある。ハイダ語も系統不明の孤立した言語である。日本語和数詞の倍数構成がほとんど崩れていないことから、和数詞は日本列島内で弥生時代初期に作られた可能性が高い。一般に語頭言の1000年での残存率は80パーセント程度と言われている。

(b) 弥生人は、中国の江南地方に住んでいた百越という民族だったという説がある。ひょっとすると、彼らが大陸から朝鮮半島と日本列島に移動し、その結果両方の地域に同じ数詞をもたらしたかもしれない。それより可能性がありそうなのは、歴史時代に日本列島から朝鮮半島に入った人が何らかの理由で日本語の数詞を残したことである。

安本美典の『日本語の誕生』や『言語の科学』に、基礎200語の語頭音の比較結果がある。それによると、日本語 (上古日本語、現代東京方言、首里方言) がいくつかの言語 (朝鮮語、中国語、インドネシア語、カンボジア語など) と統計学的に偶然とは言えない関係を持つことが示されている。このような関係は統計学とコンピュータの助けを借りて初めて検出できるのであって、従来の比較言語学の方法で

検出するのは難しい。安本は、そのほかにもビルマ系のいくつかの言語の中に日本語の身体語とよく似た単語を持つものがあることを発見している。これらの言語が本当に日本語に何らかの影響を与えたというなら、どんな人々がどの時代にどのルートで日本列島にもたらしたのだろうか。彼らの痕跡は日本の文化や遺伝子に残っているのだろうか。いずれにせよ、これらばらばらの言語との関連は、日本語の起源探求に、系統樹のモデルより流入説のモデルの方を強く支持しているように思える。

日本語の「旦那」や「僧」「袈裟」はサンスクリット語起源である。日本語の「将棋」、中国語の「象棋」、ヨーロッパの「チェス」は、インドの「チャトランガ」というゲームが元になっているそうだが、発音が何となく似ている。文化的な単語の中には、思いもかけない世界的な広がりでの借用関係があるのかもしれない。

### 3 比較言語学における数理的方法

この節は、1 節末にあげた吉田のふたつの論文から取った。ただし数学的な証明や解説は専門的になるので、相当省略した。公式がどのように使われるかを主体に解説する。

#### 3.1 ボリアの二項検定法

ハンガリーの数学者ボリアの名著に

『発見的推論そのパターン{数学における発見はいかになされるか2} 丸善 (1959)

ボリアはこの本の中で、言語同士の近さを数学的に測定する方法を提唱している。例として、前に使ったヨーロッパの言語の数詞でやってみる。英語とスウェーデン語の数詞の語頭の文字を比べてゆくと、1 と 8 で語頭の文字が異なる (1: one – en, 8: eight –

表3 数詞の語頭文字の一致数 $x_0$

	英	ス	デ	オ	ド	フ	ス	イ	ポ	ハ
英		8	8	3	4	4	4	4	3	1
ス	8		9	5	6	4	4	4	3	2
デ	8	9		4	5	4	5	5	4	2
オ	3	5	4		5	1	1	1	0	2
ド	4	6	5	5		3	3	3	2	1
フ	4	4	4	1	3		8	9	5	0
スベ	4	4	5	1	3	8		9	7	0
イ	4	4	5	1	3	9	9		6	0
ポ	3	3	4	0	2	5	7	6		0
ハ	1	2	2	2	1	0	0	0	0	
計	39	45	46	22	32	38	41	41	30	8

atta)。残りの8個の数詞では語頭の文字がすべて一致する。したがって、英語とスウェーデン語の間の一致数は8となる。このような語頭の文字の一致数をすべての言語の組について行うことによって次の表が得られる。一致数が大きければふたつの言語は近く、小さければ遠いと言える。

ふたつの文字が偶然一致する確率を $p$ とする。このとき、ふたつの言語で数詞の語頭文字の一致する回数は二項分布 $B(10, p)$ にしたがう。したがって一致数 $x$ が $x_0$ 以上である確率（上側確率）は

$$P(x_0) = \sum_{x=x_0}^{10} \binom{10}{x} p^x (1-p)^{10-x} \quad (1)$$

で計算出来る。

また、語頭文字が偶然一致する確率は

$$p = \frac{3^2 + 0^2 + 5^2 + 11^2 + \cdots + 4^2}{100^2} = 0.0948 \quad (2)$$

ここで 3, 0, 5,  $\cdots$ , 4 は、語頭文字に a, b, c,  $\cdots$ , z が現れる回数である。ただし、他の計算方法 ( $P=1/26$  など) も考えられる。

この $p$ の値を使うと、ふたつの言語の数詞の語頭文字一致数の上側確率について次の表で与えられる。

$x_0$	$P(x_0)$
0	1.000000
1	.630644
2	.243823
3	.061524
4	.010612
5	.0 <sup>2</sup> 12814
6	.0 <sup>3</sup> 10871
7	.0 <sup>5</sup> 63707
8	.0 <sup>6</sup> 24619
9	.0 <sup>8</sup> 56565
10	.0 <sup>10</sup> 58625

例えば $P(3) > 0.05$ なので、3 個の一致数では偶然の範囲を超えない。一致数が 4 だと 5 % 水準、5 だと 0.5% 水準、6 だと 0.05% 水準で、それぞれ偶然ではないと判断される。ここでは、一致数が 5 以上のとき、有意（偶然でない）と判断することにする。こうすれば、上の表により、偶然の一致なのに、間違っただけで偶然でないという可能性はほぼ 1 万回に 1 回である。これは、ほぼ 30 回連続してコインの表が出続けるほどめずらしい出来事である。4 個の一致した場合を有意としない理由は、45 組の言語の比較をしたので、4 個の一致（上側確率 0.01）だと、偶然有意になる言語の組が生じる可能性を避けるためである。

**結論。**結局、数詞の語頭文字に関しては、次のような結論が得られる。

- (a) ハンガリー語の数詞と他の言語の数詞との語頭文字の一致は偶然によるとして矛盾はない（一致数 2 以下）。
- (b) 北方系の英語、スウェーデン語、デンマーク語は強いまとまりを見せる（一致数 8 以上）。このような一致数が偶然得られる確率はきわめて小さい（有意水準 1 千万分の 5 以下）。南方系のフランス語、イタリア語、スペイン語も強いまと

まりを見せる（一致数 8 以上）。

- (c) ドイツ語は、北方系言語のグループに近い。
- (d) ポーランド語は、南方系言語のグループに近い。
- (e) オランダ語は、北方系言語のグループに近いが、南方系グループとはかなり離れ、全体的に孤立している。

従来の比較言語学によれば、英語・スウェーデン語・デンマーク語・ドイツ語・オランダ語はゲルマン系の言語であり、フランス語・イタリア語・スペイン語はラテン系の言語、ポーランド語はスラブ系の言語、ハンガリー語はウラル系の言語である。

一致数を数えるという方法は、音韻対応の法則を見出すという古典的な比較言語学の方法とは全く考え方が違う。2 番目以降の文字の情報を捨て去って、語頭の文字や音だけを比較するルーズさもある。その他諸々の理由で、専門の言語学者から批判されがちである。しかしそこでは個々の単語の比較でなく、単語のセット（例えば数詞のセット）が客観的方法で比較されるのであり、わずかの単語の選定や比較のミス、個人の恣意的判断が結論を動かすことはほとんどない。入力が同じなら、誰がやっても計算機が打ち出す出力は常に同じである。

### 3.2 オズワルトのシフト法

ポリアの検定法と同じく語頭音（または語頭文字）の一致による検定法として、オズワルトのシフト法がある。ふたつの言語 LA と LB の語彙リスト（数詞や基礎 200 語等）を用意し、それぞれの言語で意味の対応する単語には 1 から  $n$  までの番号を付けておく。同じ番号の単語はふたつの言語で同じ意味を持たなければならない。リストの単語を比べていって、語頭音の一致する組が何組あるかを数える。この一致数  $x_0$  を粗点（grosspoint）



という。

問題は、粗点 $x_0$ の大きさが偶然の一致数を超えているかどうかで、そのために偶然の一致数の分布を知る必要がある。そこでオズワルトはうまい方法を見つけた：R.L.Oswald, The detection of remote linguistic relationships, Computer Studies in the Humanities and Verval Behavior, III (1970), 117–129.

すなわち各 $i=1, \dots, n-1$ について、LAの $k$ 番目の単語とLBの $k+i(\bmod n)$ 番目の単語を $k=1, \dots, n$ まで比べ、一致数 $x_i$ を求める。LBの単語の順番をずらすとLAの単語の意味とあわないので、こうして得られた一致数 $x_1, \dots, x_{n-1}$ （背景点という）は偶然による一致数のはずである。計算を簡単にするために $x_0$ も背景点に入れておく。

コメントの追加。オズワルトも、日本でシフト法を初めて紹介した安本美典も、粗点を背景点から除いているが、理論的には入れた方が素直である。計算結果に大した差は出ない。

さらに、これら $n$ 個の背景点の平均 $m$ と分散 $s^2$ 、偏差値 $z$ 、(正規分布に関する)上側確

率 $Q_n(z)$ を求める：

$$m = \frac{1}{n}(x_0 + x_1 + \dots + x_{n-1}) \quad (3)$$

$$s^2 = \frac{1}{n}((x_0 - m)^2 + \dots + (x_{n-1} - m)^2) \quad (4)$$

$$= \frac{1}{n}(x_0^2 + x_1^2 + \dots + x_{n-1}^2) - m^2 \quad (5)$$

$$z = \frac{x_0 - m}{s} \quad (6)$$

$$Q_n(z) = \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{-t^2/2} dt \quad (7)$$

そこで、 $Q_n(z) < 0.05$ なら5パーセント水準で有意（ふたつの言語は関係あり）と判定する。

正規分布の上側確率表により、 $z$ を使って表すなら次のようになる。

- ・ $z > 1.654$ なら5パーセント水準で有意。
- ・ $z > 2.326$ なら1パーセント水準で有意。
- ・ $z > 2.576$ なら0.5パーセント水準で有意。
- ・ $z > 3.090$ なら0.1パーセント水準で有意。

安本美典がシフト法を導入したのは、日本語の起源問題の解決のためであった。安本の研究によって日本語が世界の言語の中でどのような位置を占めるかが分かったと言っても過言でない（『日本語の誕生』（大修館書店）など）。

ここでは、例として安本美典により上古日本語と中期朝鮮語の比較を紹介する。彼は日本語（上古日本語・現代日本語）と朝鮮語（中期朝鮮語・現代朝鮮・韓国語）について、200語からなる語彙リストを用意した。これは服部四郎や大野晋たちの語彙リストを使っている。多数の言語の語彙表は安本の『日本語の誕生』にある。

このような基礎語彙表は、Swadeshのリストが有名である。しかし、Swadeshのものは、代名詞や親族名称、数詞が多いなど、日本語との比較にはやや問題がある。欧米の著

項目	意味	日本語	朝鮮語
1	all	mina	motŏn
2	ash	hani	tŏi
3	bark	kana	kəptŏr
4	belly	naga	pŏi
⋮	⋮	⋮	⋮
99	woman	me	kyəptŏp
100	yellow	kŏi	nurŏ
1	all	mina	(motŏn)

図1 シフト法の原理。ひとつずつずらす。

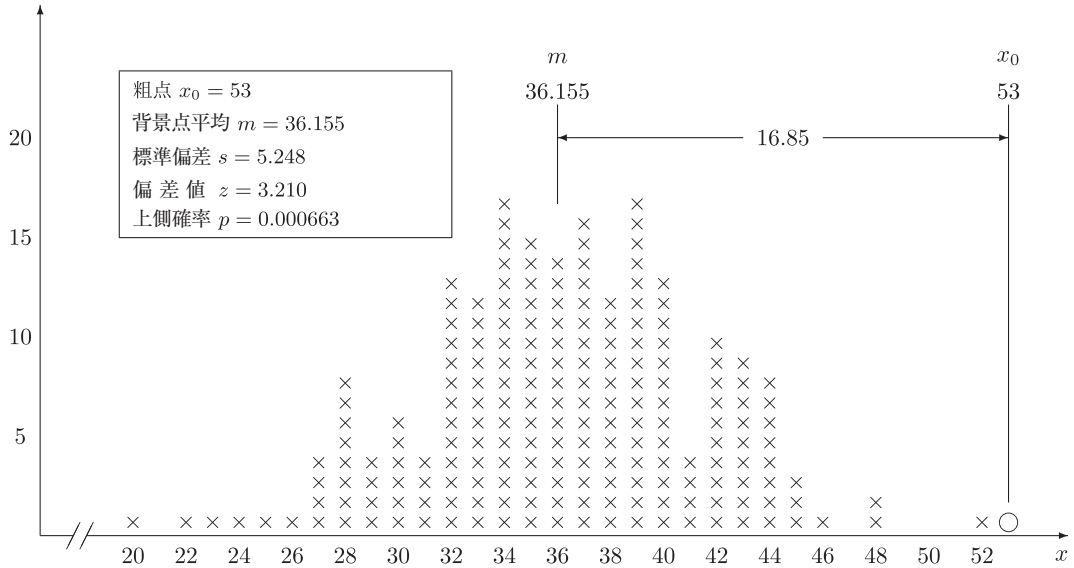


図2 「上古日本語」と「中期朝鮮語」のシフト検定(基礎200語)

名な言語学者の作成した日本語の語彙リストでも、現代日本語と古代日本語の推定形が混在するなど、日本語のネイティブ・スピーカーが作ったとは思えないものがある。

安本の使った日本語・朝鮮語・アイヌ語の語彙表は信頼性が高い。粗点を背景点に含めたことによるわずかのずれがあるものの、上古日本語と中期朝鮮語のシフト法による比較結果は当然ながら安本とほぼ同じである。安本に従い、比較は語頭子音の一致によるとし、似た音、例えば *s* 音と *t* 音をまとめた。

結論として、一致数  $x_0=53$  は平均  $m=36.155$  に比べてかなり大きい。偏差値  $z$  も 3 を超え、 $p$ -値（有意確率）は千分の 1 未満で、明白に偶然以上の一致が見られた。

オズワルトのシフト法は、音韻対応が見出だせないほど離れたふたつの言語の関係でも検出できるきわめて強力な方法である。しかし、検出力が上がった反面、個々の単語同士の類似が借用か同系かの判別や音韻対応についての情報が犠牲になっている。さらに問題点もいくつかある。

(a) 基礎200語でふたつの言語を比べるには、

全部で  $200 \times 200 = 40000$  回の単語の比較が必要となり、人手では困難である。

(b) 背景点の分布が本当に正規分布で近似できる保証がない。実際に、印欧語同士の比較で、語彙をアルファベット順に並べると、おかしい分布になる。

(c) 単語の順番によって、分布の形が変わる。

このうち手間の問題 (a) はコンピュータが解決してくれる。しかしそれを含め、すべての問題点は群論 (group theory) という抽象代数学を使って解決する。

北大理学部ホームページ『サイエンストピックス』を参照。

シフト法については、平均値  $m$  に関する次の公式 (吉田1980頃) がある。

平均値公式：

$$m = \frac{1}{n} \sum_{\lambda} a_{\lambda} b_{\lambda} \quad (8)$$

ここで  $\lambda$  は語頭子音を表す。 $a_{\lambda}, b_{\lambda}$  は、言語 KA, LB において、音  $\lambda$  で始まる単語の個数である。当然  $\sum_{\lambda} a_{\lambda} = \sum_{\lambda} b_{\lambda} = n$  である。

追加コメント。この平均値公式は二項検定法の平均値公式と同じである。これは、単語の比較を  $n$  回繰り返すとき、二項検定 (ブート

ストラップ法)では復元抽出を行い、シフト法(並べ替え検定)は、非復元抽出を行うことによる。

### 3.3 完全シフト法

サイクルの代わりにすべての順列(並べ替え)を使うシフト法を完全シフト法という。これについては背景点 $\pi$ ( $\pi$ は任意の順列)のすべての組合せモメント(とくに平均 $m$ と分散 $s^2$ )を求めることができる。

$$\begin{aligned} \text{平均 } m &= \frac{1}{n} \sum_{\lambda} a_{\lambda} b_{\lambda}. \\ \text{分散 } s^2 &= \frac{m(m+n)}{n-1} - \frac{1}{n(n-1)} \sum_{\lambda} a_{\lambda} b_{\lambda} (a_{\lambda} + b_{\lambda}) \\ \frac{1}{n!} \sum_{\pi} \binom{x(\pi)}{t} &= \frac{(n-t)!}{n!} \sum_{\Sigma t_{\lambda}=t} \prod_{\lambda} \binom{a_{\lambda}}{t_{\lambda}} \binom{b_{\lambda}}{t_{\lambda}} t_{\lambda}! \end{aligned}$$

例として、日本語と朝鮮語の音韻対応表を作る。

表4 上古日本語(J)と中期朝鮮語(K)の音韻対応表

J\K	k	m	n	p	r	t	w	y	-	
k	9	6	6	9	3	11	0	0	4	48
m	7	4	1	4	2	5	0	1	1	25
n	3	4	3	3	1	2	0	0	1	17
p	6	3	7	10	0	6	0	0	1	33
r	0	0	0	0	0	2	0	0	0	2
t	4	5	8	11	1	27	0	1	0	57
w	1	0	3	2	1	3	0	0	0	10
y	1	1	2	2	0	1	0	0	1	8
-	0	0	0	0	0	0	0	0	0	0
	31	23	30	41	8	57	0	2	8	200

偶然による一致数  $m = 36.155$ ,

偶然による一致数の標準偏差  $s = 5.1647$ ,

一致数  $x_0 = 53$ , 偏差値  $z = 3.2615$ ,

$p$ -値  $P = 0.0^3554$ , 0.1% 水準で有意。

なお、このような正方割表の対角和の検定については、コーエンのカッパ検定がある。これは正規分布の代わりに、カイ二乗分布を使うものだが、語彙数が $n=200$ と多いので、正規分布でもカイ二乗分布でも結果は変わらない。

ない。

また、両言語の音の対応として、 $t$ 音の対応  $t(J) \leftrightarrow t(K)$  が27個ある。これはカイ二乗検定ではかろうじて有意(5%水準)である。

完全シフト法については、フィッシャーの並べ替え検定と正確確率法を組み合わせた正確な $p$ -値を求める公式がある。これは ${}_2F_0$ 型の超幾何多項式を使うものである。

### 3.4 上古日本語・中期朝鮮語・アイヌ語の比較

上古日本語(J), 中期朝鮮語(K), アイヌ語(A)のシフト法による比較を行う。語彙リストは安本『日本語の誕生』にあるものを使った。一致の基準は安本と同じく次のように定義する。

$H=h=p=b=f=v=x$ ,

$t=sh=ts=d=s=z$ ,

$k=g=q=ng$ ,  $r=l$

また、対応する単語がない場合や母音だけの単語の場合‘.’で他の子音とは不一致とした。表の下段と最右列は各言語での音の分布である。

表5 上古日本語(J)とアイヌ語(A)の音韻対応表

J\A	k	m	n	p	r	t	w	y	-	
k	8	1	6	10	6	16	0	0	1	48
m	2	1	2	8	2	8	1	1	0	25
n	4	2	2	2	2	4	0	1	0	17
p	4	0	4	10	1	11	0	3	0	33
r	0	0	0	1	0	1	0	0	0	2
t	8	3	8	10	5	20	1	2	0	57
w	3	1	0	2	0	3	0	0	1	10
y	1	1	3	3	0	0	0	0	0	8
-	0	0	0	0	0	0	0	0	0	0
	30	9	25	46	16	63	2	7	2	200

3言語の各種シフト法による検定を行う。  
3言語(JKA)の語頭子音分布を $a_{\lambda}$ ,  $b_{\lambda}$ ,  $c_{\lambda}$ と

表6 アイヌ語 (A) と中期朝鮮語 (K) の音韻  
対応表

K\A	k	m	n	p	r	t	w	y	-	
k	8	2	4	5	2	10	0	0	0	31
m	3	0	4	7	2	5	2	0	0	23
n	4	0	6	6	1	11	0	1	1	30
p	5	4	4	15	3	9	0	1	0	41
r	2	0	1	1	3	1	0	0	0	8
t	5	3	6	8	5	24	0	5	1	57
w	0	0	0	0	0	0	0	0	0	0
y	1	0	0	1	0	0	0	0	0	2
-	2	0	0	3	0	0	3	0	0	8
	30	9	25	46	16	63	2	7	2	200

したとき, JAK (1) は一致数の和 $x_0$  [AJ] +  $x_0$  [AK] +  $x_0$  [KJ] に関するシフト検定である。JAK (2) は 3 言語で共通している語頭子音の個数のシフト検定である。

なお, JK のシフト検定で誤差のない正確な  $p$ -値  $P(x \geq 53)$  は, このページの最上段にある。近似値は 0.000554 である。

JK に対するシフト法の方式 (正規, 正確, 二項) の結果を比較すると以下のようになっている。

$$0.000554 (\text{正規}) < 0.00156169 (\text{正確}) \\ < 0.00238 (\text{二項})$$

正規分布を使う普通のシフト法では確率が過小に出ている。二項検定が意外と正確である。

このような状況から, 言語の比較には, ボリアの二項検定を推奨したい。

これら 3 言語の近さを正確な有意確率 ( $p$ -値) で見ると, 次のような観察ができる。

- (1) 上古日本語とアイヌ語は基礎語彙では関係が見出だせない。有意確率 20% 程度。
- (2) 上古日本語と中期朝鮮語は近い。有意確率は 1.5% 程度。
- (3) 中期朝鮮語とアイヌ語は最も近い。有意確率は 0.5% 程度。
- (4) 3 言語の比較では, 方式 1 も方式 2 でも確率は相当小さい。有意確率 1 万分の 1 未満。

アイヌ語と中期朝鮮語が最も近いのが意外である。何らかの文化的遺伝的交流があった可能性がある。

それに対し地理的に最も近い日本語とアイヌ語が近くないのも意外である。

3 言語の比較では方式 2 が有意確率が小さい。3 言語に共通の核の存在をうかがわせる。安本は, 古極東アジア語が 3 言語の核であると述べている。最近の人類学の進展もこの主張を支持していると思わせる。

### 3.5 言語年代学

時刻  $t=0$  での言語を  $L=L(0)$  であるとする。その  $t$  (千) 年後の言語を  $L(t)$  とし,  $L$  と  $L(t)$  の基礎語彙の一致数を  $x_0(t)$  とする。単語数  $n=x(0)$  である。このとき言語年代学の基本公式は

$$x_0(t) = x_0(0)r^t \quad (9)$$

表7 日本語・アイヌ語・朝鮮語の比較

	J × A	J × K	A × K	JAK(1)	JAK(2)
$x_0$	41	53	56	151	23
$m$	36.535	36.035	37.53	109.88	8.2465
$s$	5.1070	5.1635	5.2094	8.9160	2.9833
$\gamma_1$	0.0981	0.1035	0.1001	—	—
$\gamma_2$	-0.0 <sup>3</sup> 862	0.0 <sup>3</sup> 104	0.0 <sup>3</sup> 481	—	—
$z$	0.8743	3.2615	3.5455	4.6119	4.9454
$P(\text{正規})$	0.1910	0.0 <sup>3</sup> 509	0.0 <sup>3</sup> 196	0.0 <sup>5</sup> 252	0.0 <sup>6</sup> 380
$P(\text{正確})$	0.2163	0.0 <sup>2</sup> 156	0.0 <sup>3</sup> 479	—	—
$P(\text{二項})$	0.2312	0.0 <sup>2</sup> 188	0.0 <sup>3</sup> 943	0.0 <sup>4</sup> 227	0. <sup>4</sup> 104

である。

ここで、 $r$ は、千年あたりの残存率で、 $r=0.81$ 程度である。

確率過程の理論によれば、いくつかの仮定のもとで

$$x_0(t) - m(t) = (x_0(0) - m(0)) \cdot r^t \quad (10)$$

となる。ここで $m(t)$ は $L$ と $L(t)$ 間の偶然の一致数。

$m(t)=40; r=0.8$ のとき、 $x(t) \geq 50$ なら $t \leq 6.21257$  (千年)となる。

上古日本語・アイヌ語中期朝鮮語の分岐年代  
(今から何年前に分岐したか。単位千年)

$$\begin{aligned} \text{AJK} \log \frac{23 - 8.2553}{200 - 8.2553} / 3 \log 0.81 &= 4.058. \\ \text{AJ} \log \frac{41 - 36.535}{200 - 36.535} / 2 \log 0.81 &= 8.543. \\ \text{JK} \log \frac{53 - 36.035}{200 - 36.035} / 2 \log 0.81 &= 5.383. \\ \text{AK} \log \frac{56 - 37.53}{200 - 37.53} / 2 \log 0.81 &= 5.159. \end{aligned}$$

言語年代学の公式は誤差が大きいことには注意が必要である。言語年代学の基本公式は、一致数と時間との換算公式と考えるべきであろう。とくにアイヌ語と日本語の時間的距離が大きい。従来の比較言語学の方法で、日本語とアイヌ語が同系であることを証明できるとは思えない。しかし、ひょっとするとアイヌ語と韓国語については比較言語学的アプローチが可能かもしれない。

## 4 あとがき

この数年日本人の起源を扱った人類学の書籍がいくつか出版されている。

海部陽介『日本人はどこから来たのか』文藝春秋社2016

篠田謙一『DNAで語る日本人起源論』(岩波現代全書) 2015.

崎谷満『DNA・考古・言語の学際研究が示す新・日本列島史—日本人集団・日本語の成立史』, 勉誠出版2010。

崎谷満『新日本人の起源』 勉誠出版2009

どれも好著で、書評欄を飾っていた。

崎谷氏のは専門的で、とくに言語とその話者の遺伝子を合わせて論じている。篠田氏のもやや専門的で、最新の分子人類学によって日本人の起源を論じている。最後の海部氏のは読みやすく冒険心あふれる本である。

ただし日本人の起源についての考えは、重要なところで一致していない。例えば、日本人がアフリカを出てどの経路で日本列島にたどり着いたかで考えが異なる。例えば、崎谷氏はヒマラヤ北ルート、篠田氏はヒマラヤ南ルート、海部氏のは南北合流説である。

筆者の吉田は、日本語の文法が北方系であることから日本人のルートも北方系と考えてきた。篠田氏の南南説は衝撃であった。ただし後期旧石器時代人、縄文人、弥生人のルートについてはいろいろな考えがある。

海部氏の書籍には南方系の現生人類が台湾と朝鮮半島に航海技術を伝えたたとある。これで思い出したのが安本美典氏による日本語の中にインドネシア語とカンボジア語の語彙が入っているとの発見である。確かに計算してみるとかなりの語彙の流入があったと考えざるをえない。付表1, 2参照。

台湾から沖縄への航海はむずかしそうであるが、たった一人でも向こう側に付けば、その土地にはすでに人間が住んでいる。そこに自分の文化と言葉、遺伝子を残すことは不可能ではない。それに何万年という時間があるなら、その間台湾から日本列島に集団で漂着した可能性もある。

もう1つ、最近のニュースで、日本人の中に縄文人の遺伝子が15%ほど含まれているとあった。2500年で15%になったのなら、5000年では2%程度しか縄文人の遺伝子が残らな



いことになる。日本人は多様な遺伝子を持っている。その中にはかつて栄えた人たちの痕跡が残っているかもしれない。

---

**【参考文献】**

- ・ 吉田知行「言語間の距離とシフト法」数理科学1984/12。
- ・ 吉田知行「数学の応用事例—比較言語学への応用」サイエンスピックス, 北大理学部H P。
- ・ R.Oswalt: The detection of remote linguistic relationships, Computer Studies in the humanities and Verbal Behavior,III (1970)。
- ・ ポリヤ『数学における発見はいかになされるか』第2巻『発見的推論』1959。
- ・ 安本美典・野崎昭弘『言語の数理』筑摩書房1976。
- ・ 安本美典『日本語の誕生』大修館書店1983。
- ・ 安本美典『言語の科学』朝倉書店1995。
- ・ 安本美典『季刊邪馬台国』90号, 93号。
- ・ R.Gray-Q.Atkinson, Language-tree divergence times support the Anatolian theory of Indo-European origin, NATURE 426 (2003), 435-439.  
<http://www.psych.auckland.ac.nz/Psych/research/Evolution/Gray&Atkinson2003.pdf>
- ・ 崎谷満『DNAが解き明かす日本人の系譜』勉誠出版2005。
- ・ 斎藤成也『DNAから見た日本人』ちくま新書2005。

付表1 一致数と $p$ -値の表

	JPN-Tok	JPN-old	JPN-Shu	JPN-Ish	KOR-mod	KOR-mid	AINU-Ho	AINU-Sa	AINU-Ra	INDNS	ATAYAL	CAMBD
JPN-Tok	157	126	128	61	55	47	44	42	53	38	64	
	7.99e-72	7.83e-41	4.40e-43	4.98e-05	1.76e-03	7.75e-02	1.22e-01	2.12e-01	1.17e-02	2.37e-01	7.67e-04	
JPN-Old	157	121	125	58	53	41	36	34	56	42	57	
	7.99e-72	1.52e-38	2.15e-42	1.15e-04	1.88e-03	2.31e-01	4.86e-01	6.32e-01	1.05e-03	8.36e-02	5.94e-03	
JPN-Shu	126	121	124	48	43	45	43	44	52	38	50	
	7.83e-41	1.52e-38	7.63e-41	3.39e-02	1.70e-01	1.31e-01	1.41e-01	1.17e-01	8.16e-03	9.29e-02	1.89e-01	
JPN-Ish	128	125	124	54	49	48	43	38	51	38	56	
	4.40e-43	2.15e-42	7.63e-41	2.55e-03	2.86e-02	6.12e-02	1.90e-01	4.86e-01	3.21e-02	1.63e-01	2.21e-02	
KOR-mod	61	58	48	54	160	51	45	50	36	34	47	
	4.98e-05	1.15e-04	3.39e-02	2.55e-03	8.08e-80	9.55e-03	6.17e-02	8.61e-03	6.59e-01	4.01e-01	2.08e-01	
KOR-mid	55	53	43	49	160	56	48	50	39	33	50	
	1.76e-03	1.88e-03	1.70e-01	2.86e-02	8.08e-80	9.43e-04	2.20e-02	1.01e-02	4.53e-01	4.16e-01	1.01e-01	
AINU-Ho	47	41	45	48	51	56	171	140	47	35	44	
	7.75e-02	2.31e-01	1.31e-01	6.12e-02	9.55e-03	9.43e-04	1.42e-90	1.61e-54	1.56e-01	2.05e-01	6.13e-01	
AINU-Sa	44	36	43	43	45	48	171	141	48	35	41	
	1.22e-01	4.86e-01	1.41e-01	1.90e-01	6.17e-02	2.20e-02	1.42e-90	3.39e-57	1.10e-01	2.22e-01	7.23e-01	
AINU-Ra	42	34	44	38	50	50	140	141	38	36	47	
	2.12e-01	6.32e-01	1.17e-01	4.86e-01	8.61e-03	1.01e-02	1.61e-54	3.39e-57	6.19e-01	1.42e-01	2.86e-01	
INDNS	53	56	52	51	36	39	47	48	38	49	58	
	1.17e-02	1.05e-03	8.16e-03	3.21e-02	6.59e-01	4.53e-01	1.56e-01	1.10e-01	6.19e-01	5.05e-03	4.75e-02	
ATAYAL	38	42	38	38	34	33	35	35	36	49	35	
	2.37e-01	8.36e-02	9.29e-02	1.63e-01	4.01e-01	4.16e-01	2.05e-01	2.22e-01	1.42e-01	5.05e-03	4.21e-01	
CAMBD	64	57	50	56	47	50	44	41	47	58	35	
	7.67e-04	5.94e-03	1.89e-01	2.21e-02	2.08e-01	1.01e-01	6.13e-01	7.23e-01	2.86e-01	4.75e-02	4.21e-01	

日本語 (東京) × 朝鮮語 (中期, 現代), 朝鮮語 (中期, 現代) × アイヌ語 (幌別), 日本語 (東京, 上古) × カンボジア語, 日本語 (上古, 首里) × インドネシア語など

付表2 同系の単語の個数の推定 ( $x_0 - m$ ), 誤差は  $1\sigma = \sqrt{m(1 - m/n)}$  (5.0~6.3) 程度

	JPN-Tok	JPN-old	JPN-Shu	JPN-Ish	KOR-mod	KOR-mid	AINU-Ho	AINU-Sa	AINU-Ra	INDNES	ATAYAL	CAMBD
JPN-Tok	156.63	117.79	86.79	89.23	23.31	17.34	8.52	6.92	4.87	13.63	4.24	19.84
JPN-old	± 5.68	± 5.61	± 5.61	± 5.59	± 5.53	± 5.53	± 5.57	± 5.50	± 5.50	± 5.62	± 5.30	± 5.87
JPN-Shu	117.79	161.38	83.36	87.75	21.76	16.96	4.46	0.58	-1.41	18.30	7.94	15.41
JPN-Ish	± 5.61	± 5.58	± 5.53	± 5.51	± 5.45	± 5.44	± 5.46	± 5.40	± 5.40	± 5.53	± 5.32	± 5.74
	86.79	83.36	158.00	86.19	10.78	5.75	6.75	6.39	7.06	14.24	7.32	5.65
	± 5.61	± 5.53	± 5.66	± 5.54	± 5.50	± 5.51	± 5.56	± 5.47	± 5.49	± 5.53	± 5.10	± 5.87
	89.23	87.75	86.19	154.75	16.59	11.27	9.27	5.33	0.59	11.17	5.61	12.51
	± 5.59	± 5.51	± 5.54	± 5.62	± 5.51	± 5.53	± 5.59	± 5.53	± 5.51	± 5.65	± 5.21	± 5.83
KOR-mod	23.31	21.76	10.78	16.59	159.47	123.35	13.82	9.01	13.91	-1.85	1.70	5.15
KOR-mid	± 5.53	± 5.45	± 5.50	± 5.51	± 5.46	± 5.47	± 5.50	± 5.43	± 5.44	± 5.54	± 5.20	± 5.75
	17.34	16.96	5.75	11.27	123.35	155.06	18.47	11.77	13.61	1.05	1.49	7.92
	± 5.53	± 5.44	± 5.51	± 5.53	± 5.47	± 5.49	± 5.52	± 5.45	± 5.46	± 5.54	± 5.15	± 5.76
AINU-Ho	8.52	4.46	6.75	9.27	13.82	18.47	158.00	132.32	101.46	6.25	4.64	-1.28
AINU-Sa	± 5.57	± 5.46	± 5.56	± 5.59	± 5.502	± 5.52	± 5.66	± 5.59	± 5.58	± 5.70	± 5.07	± 5.92
AINU-Ra	6.92	0.58	6.39	5.33	9.01	11.77	132.32	159.01	103.61	7.52	4.35	-3.02
	± 5.50	± 5.40	± 5.47	± 5.53	± 5.43	± 5.45	± 5.59	± 5.55	± 5.51	± 5.68	± 5.09	± 5.86
	4.87	-1.41	7.06	0.59	13.91	13.61	101.46	103.61	161.69	-1.29	5.92	3.73
	± 5.50	± 5.40	± 5.49	± 5.51	± 5.44	± 5.46	± 5.58	± 5.51	± 5.51	± 5.62	± 5.05	± 5.82
INDNES	13.63	18.30	14.24	11.17	-1.85	1.05	6.25	7.52	-1.29	154.53	14.77	10.68
ATAYAL	± 5.62	± 5.53	± 5.53	± 5.65	± 5.54	± 5.54	± 5.70	± 5.68	± 5.62	± 5.88	± 5.33	± 6.01
CAMBD	4.24	7.94	7.32	5.61	1.70	1.49	4.64	4.35	5.92	14.77	162.25	1.45
	± 5.30	± 5.32	± 5.10	± 5.21	± 5.20	± 5.15	± 5.07	± 5.09	± 5.05	± 5.33	± 5.53	± 5.28
	19.84	15.41	5.65	12.51	5.15	7.92	-1.28	-3.02	3.73	10.68	1.45	144.37
	± 5.87	± 5.74	± 5.87	± 5.83	± 5.75	± 5.76	± 5.92	± 5.86	± 5.82	± 6.01	± 5.28	± 6.30

