

C-8 レーベンシュタイン距離を用いた歴史言語学の仮説評価方法の構築： スラヴ語派における印欧祖語*-eyes の反映形の問題等を例として¹

大山 祐亮 (Yūsuke Ōyama)
東京大学人文社会系研究科博士課程

要旨

歴史言語学においては、一つの語派の内部での言語変化に関する対立仮説の評価方法が確立されていない。例えば印欧祖語*-eyes に対応する共通スラヴ語の形としては*-ъje と*-i の二種類が存在しており、どちらを規則的な反映形と見なすかで論争が続いている。このような問題を解決するためには、対立仮説の優劣を評価する方法を確立することが必要である。本発表では、レーベンシュタイン距離を利用した言語の通時的変化の仮説の評価方法を提示する。まず、各仮説が想定する変化の距離を計算し、その値を説明負荷 L とする。L の値が最小のものを最も優れた仮説であると評価する。説明に必要な音変化、既知の音変化、形態変化のそれぞれの距離の総和を $lev_{a,b}(P)$ 、 $lev_{a,b}(P_K)$ 、 $lev_{a,b}(M)$ とすると、L は $L = lev_{a,b}(P) + lev_{a,b}(M) - lev_{a,b}(P_K)$ という式によって算出できる。この手法で各仮説の説明負荷を比較することで、最節約を基準とした場合の最も優れた仮説を判断できる。L の計算方法の検討を深めれば、普遍性の高い音変化仮説の評価が可能になりうる。

1. はじめに

歴史言語学は 19 世紀の時点で既に基本的な方法論が確立していた分野であるが、それ以降も方法論の改良が試みられ続けている。近年では、Nakhleh et al. (2005) などのように、歴史言語学においても最節約法や最尤法のような生物系統学の手法を応用した系統樹の作成が行われるようになってきている²。これに伴って、系統樹の形に対する仮説の比較は従来よりも少しずつ容易になりつつある。しかしながら、最節約法や最尤法といった生物系統学の手法が応用可能であるのは系統樹を描く段階までである。インド・ヨーロッパ語族バルト・スラヴ語派のように、系統関係がほぼ確定している語族あるいは語派の内部で起こった通時的変化の仮説については、現状ではまだ各対立仮説（のうち、少なくとも文献に例証されている事実と矛盾しないもの）の間の比較・評価方法が確立されていない。そのため、言語の通時的変化の詳細についてはそれぞれの研究者ごとに異なった見解が存在し、しばしばイデオロギー的な強い言葉の浴びせ合いに発展してしまう³。これを避けるために必要なのは、数値などの各仮説の優劣がはっきりとわかる形で対立仮説を評価できる手法を確立することである。

2. 従来の手法

歴史言語学の方針は、伝統的には青年文法学派の「音法則に例外なし」という原則（規則性仮説）に従い、可能な限り（形態変化ではなく）音変化によって説明することを試みるというものであった。例えば、所謂グリムの法則 (Grimm 1822)⁴ の場合、印欧祖語 (PIE) *d の位置に現れるゲルマン祖語*t の例の全てを形態変化によって説明する仮説よりは、PIE *d > ゲルマン祖語 *t⁵ という規則的な音変化が存在するとする仮説の

¹ 謝辞：本研究は JSPS 科研費 19J22945 の助成を受けたものである。また、本発表で用いられるスラヴ語の音を表記する文字のうち、注意を要するものは以下の通りである。まず、共通スラヴ語の*ъ および*ь はそれぞれ極めて短い i、u の音、*y は中舌母音[i]、*ǫ は o に対応する鼻母音であったと推定される。

² この分野には多くの研究の蓄積が存在するが、紙幅の都合のため省略する。なお、本発表の内容は生物系統学を利用して系統樹を描くというものは根本的に別物である。使用している概念は類似しているが、目的と手法は異なっている。

³ 直近のものとしては、"Attempts to solve classic problems in terms of dominance patterns have resulted in complete failure." (Kortlandt 2018: 295) という批判に対して、"However, it is quite the opposite – mainstream Western Balto-Slavic historical linguistics has been sleeping on the problem and is still desperately clinging on to implausible sound laws and a Vedic-centric Proto-Indo-European reconstruction, which has been abandoned long ago in other spheres of Indo-European linguistics." (Kapović 2019: 120)[下線は発表者による]という反論が行われた例が挙げられる。

⁴ 詳細については例えば Collinge (1985: 63–76) 等を参照せよ。

⁵ 本発表においては、>で音変化、→で形態変化をそれぞれ示す。

方が優れているということが出来る。しかし、この方針で各仮説間に明確な優劣をつけられるのは、規則的な音変化のみを必要とする仮説と、一つ以上の形態変化を必要とする仮説が対立する場合に限られる。このような状況は実際の歴史言語学においてはむしろ稀であり、音変化のみでは説明不可能な問題にしばしば直面する。スラヴ語歴史言語学の例を一つ挙げると、共通スラヴ語 (LCS) における PIE *-eyes の反映形は、以下のように二種類に分化してしまうことが知られている。

例：PIE *pont-h₁/h₂-ey-es > LCS *pqt-**lje** 「道 (M.NOM.PL.)」⁶

PIE *nok^wt-ey-es > LCS *notj-**i** 「夜 (F.NOM.PL.)」

このような場合、規則的な音変化によって説明できるのは最大でもどちらか一方のみであり、片方もしくは両方が形態的变化を受けている可能性がある。この問題について、先行研究は PIE *-eyes > LCS *-lje が規則的な音変化だと主張するもの (e.g. Olander 2015: 64, 66) と、PIE *-eyes > LCS *-i が規則的な音変化だと主張するもの (e.g. Hill 2016: 216–218) に分かれて論争が続いており、未解決のままである。

この問題を解決困難なものとしているのは、PIE *-eye-の再建される形態素の例が乏しいということと、音韻論的にはこれらの音変化の両方が十分あり得るということである。その帰結として、どちらの仮説を採用する立場からも決定的な証拠を提示することができず、意見の一致をみることが事実上不可能となっている。このことから、最節約法や規則性仮説といった既存の手法に加えて、各仮説間の優劣を判定する新しい方法を導入することが必要であるということがわかる。

3. レーベンシュタイン距離の歴史言語学的な重み付け

本論文では、**レーベンシュタイン距離** (編集距離、Levenshtein 1965) の計算方法を利用して、言語の通時的変化の仮説の評価方法を構築し、提示する。レーベンシュタイン距離とは、ある文字列を別の文字列に変換するのにどれほどの操作が必要であるのかという指標である。例えば、*lake* を *take* に変換するためには、*l* を *t* に置換する (あるいは、*l* を削除してから *t* を挿入する) ことが必要である。通常、この両者のレーベンシュタイン距離は、置換という操作を認めるならば置換一回分の 1、認めないならば削除と挿入を一回ずつ行うため 2 となる。この概念は一般的には自然言語処理の分野で用いられるが、近年では方言間の差異の大きさを測る指標として用いられ (Kessler 1995)、『百鬼夜行絵巻』のような模写によって描き継がれる写本の系統推定 (山田 2012) などに応用されたりしている。

『百鬼夜行絵巻』のような写本の系統推定の場合、二つの写本間のレーベンシュタイン距離が決まれば、レーベンシュタイン距離によって想定されるのと同じ操作が模写の過程で起こっていると仮定してよい (山田 2012: 21)。例えば、真珠庵本と歴博 A 本という二伝本の間で二箇所 of 図像配列の違いがある (即ち、レーベンシュタイン距離が 2 である) ならば、模写の過程で二箇所の配列が入れ替わったと見た目の通りに解釈して問題が生じない。

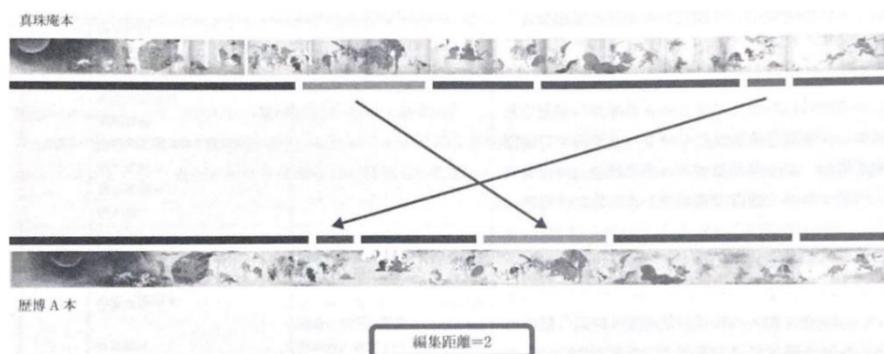


図 1：『百鬼夜行絵巻』におけるレーベンシュタイン距離の応用の例 (山田 2012: 24)

⁶ これらの例はスラヴ語の形に合うように印欧祖語の形態素を組み合わせたものにすぎないことを付言する。

しかしながら、歴史言語学の場合には、写本系統の場合とは異なり、レーベンシュタイン距離をそのままの形で利用することは不可能である。まず、置換・削除・挿入のような操作の実行方法に、音変化と形態変化という二種類がある。例えば、PIE *-os が LCS *-b̥ になるという変化があった場合、その変化の経路として、PIE *-os > LCS *-b̥ という音変化を想定するのと、PIE *-os → LCS *-b̥ という形態変化を想定するのでは本質的に別の仮説となる。このため、両者を同列に扱うかどうかをまず検討する必要がある。音変化についても、k > g のような起こりやすいものと、t > e のようなまず起こらないものを同列に扱うことはできない。

以上の要因により、歴史言語学においては、二つの文字列を比べて単純にレーベンシュタイン距離を算出し、その通りの操作を想定することによって言語の通時的な変化を説明する、ということとは不可能である場合の方が多い。したがって、レーベンシュタイン距離を歴史言語学の議論に応用できるようにするためには、何らかの重み付けをすることが必要である。

重みづけに際して、まず、任意の要素の削除と挿入の距離をそれぞれ 1.00 と設定する。モーラあるいは音調の変化や、挿入母音などがこれに該当する⁷。そして形態的変化のレーベンシュタイン距離を、一つの形態素におこる一つの変化につき 1.00 と設定する⁸。そのうえで、音変化のうち音の弁別素性のみが変化するもの (i.e. 置換) について、Fontan et al. (2016) の導入した PWLD (phonologically weighted Levenshtein distances) の定義を用いる。これは、その言語の内部で対立する弁別素性⁹のうちのいくつが変化するのか、という基準によってレーベンシュタイン距離を算出するものである。例えば、母音について [±back]、[±high]、[±low] が弁別的である言語において e > i という変化が起こった場合、これは [-high] から [+high] という一素性のみの変化であるため、レーベンシュタイン距離は 0.33 (= 1/3) である。ただし、隣接した音からの素性の受け渡しが存在する場合、すなわち同化が起こった場合は 0 と計算するものとする。すなわち、それまでその周辺に存在していなかった素性の値が生じる場合に、距離が生じることとなる。このような重み付けをすることで、レーベンシュタイン距離を歴史言語学に応用することが可能となる。

4. レーベンシュタイン距離を用いた歴史言語学の仮説の評価方法

以上の重み付けを行ったレーベンシュタイン距離の計算方法を用いて、新しい仮説評価方法を導入する。

まず、各対立仮説ごとにその仮説が想定する変化のレーベンシュタイン距離¹⁰を計算し、算出された値をその仮説の説明負荷 L とする。そして、最節約原理 (parsimony) に基づき、 L の値が最も小さいものを最も優れた仮説であると評価することとする¹¹。計算の際の注意点は以下の通りである。まず、説明しようと試みる形態素以外の部分において既に知られている規則的な音変化は、どの対立仮説においてもその存在が前提とされるため、対立仮説間の比較をする際にはその分の値を差し引いて考える必要がある。また、音変化の定義は例外が存在しない変化であることを原則とし、散発的でありうるのは音位転換と異化のみとする¹²。形態変化の定義は動機の説明を必須とする。

以上のことを踏まえると、説明に必要な音変化の距離の総和、既知の音変化の距離の総和、そして説明に必要な形態変化の距離の総和をそれぞれ $lev_{a,b}(P)$ 、 $lev_{a,b}(P_K)$ 、 $lev_{a,b}(M)$ とすると、レーベンシュタイン距離をもとにした言語変化仮説の説明負荷 L は以下のような式によって算出することができる。

$$L = lev_{a,b}(P) + lev_{a,b}(M) - lev_{a,b}(P_K)$$

⁷ なお、代償延長の場合にはモーラの削除も挿入も起きないため、脱落する分節音の距離 1.00 のみを加算する。

⁸ 音変化が形態的条件によって妨げられる場合にも、一つの形態素につき 1.00 と定義する。

⁹ なお、変化の前後で音素の数が増える場合、変化の前後の弁別的な素性の数を比較し、多い方を用いることとする。

¹⁰ すなわち、本論文の手法で用いるレーベンシュタイン距離は、仮説の前提条件に応じて実際には通り得ない経路が存在し、かつ同じ経路でも通り方が二種類あるという点で、自然言語処理等の他の文脈で用いられる定義とは異なる。

¹¹ なお、ベイズ法などの確率に基づくモデルでなく最節約原理を利用する理由としては、音変化だけでなく形態変化の可能性も含めて扱う場合、形態変化は変化前と変化後が必ずしも似た音にならないため、変化後の形の候補が理論上無限に存在し、変化の確率を計算するのが現実的ではないということが挙げられる。また、確率に基づくモデルに対する発表者の見解は Pereltsvaig and Lewis (2015) や濱田 (2019) に近い。

¹² これらの変化はしばしば散発的に起こることが知られている (e.g. Brugmann and Osthoff 1878: xiv)。

これは端的に言って、その変化の説明のために新しく定義する音変化と形態変化の距離の合計である。なお、この式には既知の音変化の距離を加えてから引くという操作が含まれているが、これは二度手間である。したがって、実際の運用においては、最初から既知の音変化の距離を計算に含めずに、新しく定義する音・形態変化の距離のみの和を算出してしまっても支障がない。このようにして算出した説明負荷 L の値が最小となる仮説が、最節約原理に基づいた場合の最も妥当な仮説であるといえる。¹³

以下ではこの手法の応用例を提示するが、その前にまず、古ロシア語 (Old Ru.) $-ě$ (狭い e : (Borkovskij and Kuznecov 2009: 135) の現代ロシア語 (Ru.) における反映形、という文献中で例証されている通時的変化の問題を例として、この手法が正しい結果を予測することができるか検証する。Old Ru. $-ě$ をもつ形態素の Ru. における反映形は以下の通りである。なお、語中では Old Ru. $ě > Ru. e$ が既知の変化であり (e.g. Old Ru. *cěna* $>$ Ru. *cena* 「値段」)、実際には語末でも同様に Old Ru. $-ě > Ru. -e$ という変化が起こっていることが知られている (Matthews 1960: 155, 162, 165)。

- (1) o, \bar{a} 語幹名詞の単数前置格語尾 Old Ru. $-ě > Ru. -e$ (e.g. Old Ru. *gorodě* $>$ Ru. *gorode* 「街で」)
- (2) \bar{a} 語幹女性名詞の単数与格語尾 Old Ru. $-ě > Ru. -e$ (e.g. Old Ru. *ženě* $>$ Ru. *žene* 「妻に」)
- (3) 形容詞長語尾形の複数主・対格語尾 Old Ru. $-ě > Ru. -e$ (e.g. Old Ru. *gorodě* $>$ Ru. *gorode* 「街」)
- (4) 一・二人称代名詞の単数与・前置格語尾 Old Ru. $-ě > Ru. -e$ (e.g. Old Ru. *tobě* $>$ Ru. *tebe* 「私に」)
- (5) *ona* 「彼女」の単数生・対格形 Old Ru. $eě > Ru. ee$
- (6) $y\bar{a}$ 語幹名詞の単数生格語尾 Old Ru. $-ě > Ru. -i$ (e.g. Old Ru. *zemlě* $>$ Ru. *zemli* 「大地の、国の」)
- (7) $y\bar{a}$ 語幹名詞の複数主・対格語尾 Old Ru. $-ě > Ru. -i$ (e.g. Old Ru. *zemlě* $>$ Ru. *zemli* 「国々」)

この変化の過程が文献中で例証されていないと仮定した場合、この問題に対しては、語中と同じ Old Ru. $-ě > Ru. -e$ か、あるいは語末独自の Old Ru. $-ě > Ru. -i$ という音変化による二通りの説明が考え得る。

まず、前者が規則的な変化だと仮定した場合には、Ru. $-e$ の例は語中と同じ音変化によって説明することができるため¹⁴、Ru. $-i$ を示す二例について形態的变化を仮定すればよい。したがって、 $L = 2.00$ となる。一方で、後者が規則的な変化だと仮定した場合には、Old Ru. $-ě > Ru. -i$ を新しい音変化として定義しなければならない。Old Ru. においては $[\pm\text{back}]$ 、 $[\pm\text{high}]$ 、 $[\pm\text{low}]$ 、 $[\pm\text{round}]$ の四種が弁別的であるため、 $[\text{high}]$ の値のみが変化するこの変化の説明負荷は 0.25 ($= 1/4$) である。これに加えて $-e$ を示す五例について形態的变化を仮定しなければならないため、 $L = 0.25 + 5.00 = 5.25$ である。したがって、Old Ru. $-ě > Ru. -e$ が規則的な変化だと仮定する方が最節約原理の観点からみてよい仮説であると結論づけられる。これは文献で例証されている変化と同一であるため、この手法で仮説を正しく評価できることが示された。

次に、この手法を上述の PIE $*\text{-eyes}$ の例に応用して、先行研究の二つの仮説の説明負荷 L を計算する。なお、共通スラヴ語の前段階 (印欧祖語の喉音の脱落后の状況、あるいはバルト・スラヴ祖語) の母音体系は $*a, *e, *i, *o, *u$ の五母音とその長短であり、 $[\pm\text{back}]$ 、 $[\pm\text{high}]$ 、 $[\pm\text{low}]$ の三種が弁別的であったと推定される。まず、PIE $*\text{-eyes} > \text{LCS } *-\text{ьje}$ が規則的な音変化であると仮定した場合、この変化のレーベンシュタイン距離は、 $*e > *ь / _je$ という $[\text{high}]$ の値の変化による 0.33 と、形態変化ひとつの距離 1.00 の和であるため¹⁵、 $L = 1.33$ となる。一方で PIE $*\text{-eyes} > \text{LCS } *-\text{i}$ が規則的な音変化であると仮定した場合には、 $*ey$ の直後での $*e$ の脱落¹⁶

¹³ なお、任意の形態あるいは音素について何らかの中間段階を想定することが必要となる場合には、中間段階に至るまでの変化とそれ以降の変化に仮説を分割して、それぞれについて L の計算を行う。また、和をとった時点で、数学的な意味での距離とは乖離している。すなわち、ここでの L はレーベンシュタイン距離の理念を利用しているものの、数学的な距離とは別物である。この点について、塚越柚季氏 (p.c.) に、距離ではなくベクトルにして計算すればより数学と親和性の高い分析が可能ではないかというご指摘を頂戴した。これは今後の検討課題としたい。

¹⁴ 厳密に言えば、Old Ru. $-ě > Ru. -e$ という変化に位置の制約がないと仮定することが必要となる。このような仮定で説明することが可能である場合には既知の変化に含めてよいこととする。なお、この変化を既知と仮定しなくてもこの例の結論は変化しない。

¹⁵ PIE $*\text{-s} > \text{LCS } \emptyset$ は既知の音変化であり、特に異論の余地がない (e.g. PIE $*\text{meiHros} > \text{LCS } *mirь$ 「平和」)。

¹⁶ PIE $*\text{ej} > *i\bar{i} > *i > \text{LCS } *i$ も考えうるが、この場合には $L = 1.67$ であり、 $*e$ の脱落を想定する方が L の値が小さい。

による距離 1.00 と形態変化の 1.00 を足して $L = 2.00$ となるため¹⁷、PIE **-eyes* > LCS **-bje* を規則的な音変化とみなす方が最節約の観点からみて優れた仮説であると判断できる。このように、対立仮説の説明負荷を比較することで、最節約を基準とした場合の最も優れた仮説を判断することが可能となる。

5. スラヴ語歴史言語学への応用例

以下では、スラヴ語歴史言語学において未だ決着がつかない問題をさらに二つ取り上げて、本発表で提案した手法の応用例を提示する。

5.1 語源的に期待されない位置に現れる鼻音（形態変化による説明が妥当となる場合）

共通スラヴ語で鼻母音が現れるにもかかわらず、印欧祖語に対応する鼻音が存在しない、という屈折語尾がスラヴ語派には二つ存在する。この問題について、通常は両方の形態素について鼻音 **-m* の添加が起こったと説明される (e.g. Olander 2015: 163–166, 307–309)。一方で、先行研究の一部には、これらについて **-m* ではなく **-ō-mi* や **-ā-mi* といった中間段階を再建して、そこから **i > Ø / V̄m_#* という規則的な音変化によって語末の *i* 音が脱落し (所謂 *i-Apocope*)、それによって生じた **-Vm* が鼻母音となる、という説明を行うものが存在する (e.g. Vaillant 1950: 89)。この仮説によって説明される形態は以下の通りである。

(i) Thematic の動詞の現在一人称単数形語尾 **-q*

例：PIE **b^her-o-H* > **berō* > **berō-mi* > **berōm* > LCS **berq* 「運ぶ (prs.1sg.)」(?)

(ii) 人称代名詞および(y)ā 語幹女性名詞の単数具格語尾 **-q*

例：PIE **g^wen-ch₂-h₁* > **g^wenā* > **genā-mi* > **genām* > LCS **ženq* 「女 (instr.sg.)」(?)

しかしながら、この *i-Apocope* には、Athematic の動詞の現在一人称単数形語尾 (e.g. LCS **jьma-мь* 「持つ」) や、条件法助動詞一人称単数形 LCS **bi-мь* (< **bī-mi* < **b^huh₂-iH-mi*) という例外が存在する。さらに、系統的に最も近いバルト語派の語形との対応を考慮すると、(ii)はバルト語派とスラヴ語派の共通改新であるが、(i)はスラヴ語派独自の改新として定義せざるを得ない (Olander 2015: 308)¹⁸。したがって、この仮説は *i-Apocope* が 2 回と、例外の Athematic の動詞の一人称単数形に形態変化を想定するため、 $L = 2.00 + 1.00 = 3.00$ (条件法助動詞の *bimь* を別に計算すると $L = 4.00$) となる。一方で、旧来の説明では、両形態素に **-mi* ではなく接辞 **-m* の添加という形態変化を想定するため、形態変化二つにより妥当度 $L = 2.00$ となる。したがって、*i-Apocope* よりも形態変化を想定する方が妥当であると結論づけられる。

5.2 二人称複数代名詞の由来（音変化による説明が妥当となる場合）

共通スラヴ語の二人称代名詞の複数主格 **vy* の由来については、以下の三種類の対立仮説がある。

(a) PIE **yuH(s)* > **wōs* > **wōns* > LCS **vy* (Vondrák 1908: 72–73 ほか)

(b) PIE **yuH(s)* > **wōs* > LCS **vy* (Meillet 1934: 454)

(c) PIE **yuH(s)* > **wū(s)* > LCS **vy* (Reinhart 2002: 138 ほか)

既知の **-ō(n)s* > LCS **-y* (e.g. Meillet 1934: 130–131) という変化を除くと、(a)は形態変化二つのため $L = 2.00$ 、(b)は形態変化一つのため $L = 1.00$ である。(c)は、頭子音の **y > *w* の類推の元となる形が存在しないため形態変化としては成立せず、音変化とみると例外が最低でも二つ (LCS **jigo* 「くびき」 < **jūgān*、**šiti* 「縫う」 < **šjūtěj*) 存在するため、 $L \geq 2.00$ となる。したがって、これらの仮説は説明負荷が低いものから順に(b)、(a)、

¹⁷ PIE **e_i* > LCS **i* は既知の音変化であり、特に異論の余地がない (e.g. PIE **meiHros* > LCS **mirь* 「平和」)。

¹⁸ この場合、Athematic の動詞の語尾 **-mi* はこの二回の変化の両方の形態的な条件による例外となってそのままの形が保たれたことになる。これは考え難いため、この点を根拠とすれば本発表の手法を用いずとも *i-Apocope* 説に反駁することは可能である。

(c)という順になる。このように、本発表の手法を用いると各対立仮説間の優劣が明確になる。

6. 本手法の利点・意義

この手法を導入することによる利点は三つ挙げられる。まず一つ目は、説明負荷が数値の形で算出できるため、対立仮説間の比較がやりやすく、優劣がわかりやすいという点である。

二つ目は、明らかに冗長と思われる仮説を排除する根拠を与えられるということである。例えば、スラヴ語派における s アオリストの一人称単数形 (e.g. PIE *wēdʰ-s-ŋ > LCS *věsь 「導く」) は、既知の音変化のみを用いて (i.e. $L = 0.00$)、PIE *-sŋ > *-sum > LCS *-sь と説明することが可能であるが、これに対して Meillet (1934: 323) などのように PIE *-sŋ → *-som > *-sum > LCS *-sь と形態変化を一つ増やすと $L = 1.00$ となり、冗長性が L の値の差としてはっきりと現れる。この数値の差を冗長な仮説を棄却するための反論しにくい根拠として用いることができる。このことにより、音変化が提案される、論敵によって反例が指摘される、指摘された例を半ばアドホックな形態変化の定義によって説明して自身の仮説を延命させる、という歴史言語学の論争でしばしば発生する状況を回避することができる。

三つ目は、この手法による対立仮説間の比較が、複数の仮説が対立して膠着状態に陥ってしまった状況を打破するきっかけとなるという点である。この手法で暫定的に対立仮説間に序列を作ってしまうことによって、論争の焦点を、対立仮説の主張する音変化に例外が存在しないか、形態的变化に妥当な動機が存在するか、この説明負荷 L の計算方法に改善余地はないか、という三つに絞ることが可能となる。

7. 今後の課題

現時点では、この手法にも少なくとも四つの課題ないし注意点が存在する。一つ目は、二重母音化のような要素の数が変わる場合の計算方法が未定義であるという点である。ö を /oo/ と解釈する等の方法が正当化されるかは今後の課題としたい。少なくとも現時点では、計算方法が定義済みである変化が起こっている例にのみこの手法を応用するのが妥当である。二つ目は、あまりにも音が大きく変化してしまっているものは、計算が難しいという点である (e.g. PIE *dw- > Arm. erk-: Meillet 1894)。三つ目は、その言語でどのような弁別素性の対立があるのか、そしてどの音変化が既知なのかという判断が各研究者の能力と倫理観に依存しているという点である。この手法に限らず、尤度などの計算を要する手法全般に、p 値ハッキングに代表されるような不正や¹⁹、所謂 GIGO (Garbage in, garbage out) の危険性があることには注意すべきである。四つ目は、学派毎に異なった道具立てを用いている場合にはこの手法の応用が困難であるという点である。超分節素に関する分析がその最も顕著な例として挙げられる²⁰。例えば、バルト・スラヴ語アクセント論は、学派毎の分析方法の差異が大きい。生成音韻論の研究では自律分節素などの通言語的とされる概念を用いるが、モスクワ学派では±の *valency*、オランダ学派は *acute* と *circumflex* という独自の概念をそれぞれ用いる。これらの理論を同じ基準で測ることは困難である。したがって、この場合にはそれぞれの学派の道具立てが正当なものであるかをまず議論の対象とする必要がある。

8. 結論

本発表では、従来の歴史言語学の方法論に基づいて提起された様々な対立仮説を比較し、評価するための方法を提起した。これによって、これまでの歴史言語学の方法論で掘り下げ尽くした問題をさらに掘り下げるための、あるいは先行研究で議論が紛糾した状態を解決するための道具が一つ追加されたということになる。しかしながら、音変化が例外のないものとして成立するのか、類推に動機が存在するのかというような議論については従来の歴史言語学の方法論に則って行うこととなる。したがって、この手法を採用したから

¹⁹ ただし、現代の印欧語比較言語学では基本的に全ての研究者が同じデータを見ながら研究を行っているため、心理学のような他の分野ほどこの危険性は高くはないと思われる。例えば印欧語において、自らの仮説に不利なデータを隠してこの手法の計算を行った場合、他の研究者はその隠されたデータの知識も有しているため、不正を指摘することが可能であると考えられる。

²⁰ なお、音調に関してレーベンシュタイン距離を設定すること自体は不可能ではない (e.g. Yang and Castro 2008)。

とって歴史言語学の方法論が大きく変わるわけではない。

また、この手法が応用可能であるのは、本発表中のロシア語の例のように例証されている変化か、印欧語族のように主要な音変化が既に確立されている語族の場合に限られる。まだ研究があまり進んでいない語族の場合には、伝統的な比較法や、生物系統学的手法としての最節約法等を利用する方が有益である。応用可能な手法の数が増えれば増えるほど、どのような状況でどのような手法を用いるべきかということを検討し、各種の手法の適切な取捨選択を行う必要性は増大する。方法論が進歩すればするほど、方法論に対する理解が重要となるであろう。

略語

Arm.: Armenian / LCS: Late Common Slavonic / PWLD: phonologically weighted Levenshtein distances / Ru.: Russian

参考文献

- Borkovskij, V.I. and Kuznecov, P.S. (2009) *Istoričeskaja grammatika russkogo jazyka*. 5th ed., Moskva: Knižnyj Dom LIBROKOM.
- Brugmann, Karl and Hermann Osthoff (1878) Vorwort. *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen* 1: III–XVI.
- Collinge, Neville E. (1985) *The laws of Indo-European*. (Amsterdam Studies in the Theory and History of Linguistic Science, series IV, 35.) Amsterdam / Philadelphia: J. Benjamins.
- Fontan et al. Fontan, Lionel, Ferrané, Isabelle, Farinas, Jérôme, Pinquier, Julien and Xavier Aumont (2016) *Using Phonologically Weighted Levenshtein Distances for the Prediction of Microscopic Intelligibility*. Annual conference Interspeech (INTERSPEECH 2016), Sep 2016, San Francisco, CA, United States. 650–654. hal-01474904f
- Grimm, Jakob (1822) *Deutsche grammatik*. Band I. 2nd ed. Göttingen: Dieterich.
- 濱田武志(2019)『中国方言系統論：漢語系諸語の分岐と粵語の成立』東京：東京大学出版会。
- Hill, Eugen (2016) Phonological evidence for a Proto-Baltic stage in the evolution of East and West Baltic. *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 13: 205–232.
- Jasanoff, Jay H. (1983) A rule of final syllables in Slavic. *Journal of Indo-European Studies* 11: 139–149.
- Kessler, Brett (1995) Computational dialectology in Irish Gaelic. In: *Proceedings of the Seventh Conference on European Chapter of the Association for Computational Linguistics*, 60–67. San Francisco: Morgan Kaufmann.
- Levenštejn, Vladimir Iosifovič (1965) Dvoičnyje kody s ispravleniem vypadenij, vstavok i zameščenij simvolov. *Doklady Akademij Nauk SSSR* 163(4): 845–848.
- Kapović, Mate (2019) Shortening, Lengthening, and Reconstruction: Notes on Historical Slavic Accentology. *Rasprave Instituta za hrvatski jezik i jezikoslovlje* 45(1): 75–133.
- Kortlandt, Frederik (1983) On final syllables in Slavic. *Journal of Indo-European Studies* 11: 167–185.
- Kortlandt, Frederik (2018) Rise and Fall of Vowel Length in Slavic. *Rasprave Instituta za hrvatski jezik i jezikoslovlje* 44(1) 287–297.
- Matthews, William Kleesmann (1960) *Russian historical grammar*. London: University of London, Athlone Press.
- Meillet, Antoine (1894) Notes arméniennes. *Mémoires de la Société de Linguistique de Paris* 8: 153–165.
- Meillet, Antoine (1934) *Le slave commun*. 2nd ed. Paris: Champion.
- Nakhleh, Luay, Ringe, Donald and Tandy Warnow (2005) Perfect Phylogenetic Networks: A New Methodology for Reconstructing the Evolutionary History of Natural Languages. *Language* 81(2): 382–420.
- Olander, Thomas (2012) Proto-Indo-European *-os in Slavic. *Russian Linguistics* 36: 319–341.
- Olander, Thomas (2015) *Proto-Slavic Inflectional Morphology: A Comparative Handbook*. Leiden: Brill.
- 大山祐亮(2020)「共通スラヴ語における印欧祖語*-osの反映形：屈折体系の変化に着目した説明」日本言語学会 160 回大会口頭発表。オンライン（言語学会 HP），2020 年 6 月 20 日。[予稿集：7–13]
- Pereltsvaig, Asya and Martin W. Lewis (2015) *Indo-European Controversy: Facts and Fallacies in Historical Linguistics*. Cambridge: Cambridge University Press.
- Reinhart, Johannes (2002) Morphologische Innovationen des Altkirchenslavischen. *Wiener Slavistisches Jahrbuch* 48: 133–148.
- Schmalstieg, William (1983) Slavic *Kamy* and *Mati*: A Structural View. *Journal of Indo-European Studies* 11: 151.
- Vaillant, André (1950) *Grammaire comparée des langues slaves*. Tome 1: Phonétique. Paris and Lyon: IAC.
- Vondrák, Václav (1908) *Vergleichende Slavische Grammatik*. vol. 2. Formenlehre und Syntax. Göttingen: Vandenhoeck und Ruprecht.
- Yang, Cathryn and Andy Castro (2008) Representing Tone in Levenshtein Distance. *International Journal of Humanities and Arts Computing* 2(1–2): 205–219.
- 山田奨治(2012)「『百鬼夜行絵巻』写本の系統」中尾央・三中信宏編『文化系統学への招待：文化の進化パターンを探る』17–34. 勁草書房。