
УДК 81-115

Валентина Перебийніс (м. Київ)

СТУПЕНІ ГЕНЕТИЧНОЇ СПОРІДНЕНОСТІ УКРАЇНСЬКОЇ МОВИ З ІНШИМИ СЛОВ'ЯНСЬКИМИ МОВАМИ: СТАТИСТИЧНИЙ ПІДХІД

У статті описано статистичний підхід до визначення ступенів генетичної спорідненості слов'янських мов шляхом попарного зіставлення мов за диференційними ознаками, що відображають фонетичні закони історичного розвитку слов'янських мов. Якщо певна диференційна ознака властива обом порівнюваним мовам, то вона є об'єднувальною, і роз'єднувальною, якщо властива лише одній із зіставлюваних мов. Що більше об'єднувальних ознак мають дві зіставлювані мови, то вищим є ступінь їх спорідненості.

Ключові слова: слов'янські мови, генетична спорідненість, об'єднувальна диференційна ознака, роз'єднувальна диференційна ознака, статистичний підхід, таблиця спряжених ознак.

Визначення генетичної спорідненості мов — завдання порівняльно-історичного мовознавства. Виходячи з твердження про походження кожної мови з єдиного джерела, або прамови, ця галузь мовознавства зосереджується на дослідженні тих змін, які відбуваються в діалектах певної прамови і приводять до розщеплення кожного з них на самостійні мови. Ці зміни стосуються фонетичної, граматичної й лексичної систем і, якщо вони регулярні, приводять до встановлення законів розвитку мов. Зіставлення мов за цими законами, тобто регулярними змінами, має результатом визначення генетичної спорідненості мов і побудову генеалогічного дерева мов, виокремлення мовних сімей, груп і підгруп. Як твердять мовознавці, які виконують такі дослідження, найточніші результати дають дослідження фонетичних законів, на яких, власне, базується порівняльно-історичний метод.

Але якими б ефективними не були фонетичні закони, їх дія не є абсолютно рівномірною, тому існують різні ступені спорідненості мов, визначені за фонетичними законами. Встановити ці ступені можна за допомогою статистичних методів. Одним із них є визначення спряженості ознак, або коефіцієнта подібності [3: 123-125]. В основі цього методу лежить зіставлення диференційних ознак, за якими порівнюють мови. Схема його застосування така:

© В.І. ПЕРЕБИЙНІС, 2012

ISSN 1682-3540. Українська мова, 2012, № 3

31

Таблиця 1. Співвідношення диференційних ознак у двох мовах.

| | | | |
|----------|------------|------------|----------------|
| 2-а мова | + | – | Разом |
| 1-а мова | | | |
| + | a | b | a+b |
| – | c | d | c+d |
| Разом | a+c | b+d | a+b+c+d |

1. Укладають список диференційних ознак (ДО), наприклад, фонетичних законів. Їх використання є доцільним тому, що вони досить формалізовані, і їх визначення не становить труднощів. А відомо, що точне визначення одиниць підрахунку — одна з основних вимог до організації статистичних досліджень.

При цьому кожна ДО повинна бути однозначною, тобто відображати тільки одну зміну. Якщо звук прамови дає різні рефлексі, тобто, змінюється по-різному в різних умовах історичного розвитку, встановлюється кілька ДО. Наприклад, прадавній звук ***ě** дає рефлексі **i**, **e**, **a**, **o** (див. стор. 33), отже, встановлюється чотири ДО, за якими розрізняються мови.

2. Встановлюють список мов, які будуть досліджувати.

3. Кожну з відібраних мов зіставляють зі списком ДО, щоб встановити, які ДО властиві мові, а які не властиві. Будують таблицю, у верхньому рядку якої перераховують мови. Якщо ДО властива мові, то на перетині відповідного стовпчика й рядка ставлять +, а якщо не властива, ставлять – (Табл. 2).

Таблиця 2. Наявність чи відсутність ДО в зіставлених мовах.

| ДО | Укр. | Рос. | Блр. | Чеш. | Слц. | Плс. | Блг. | Срх. | Слв. |
|----------------|------|------|------|------|------|------|------|------|------|
| * ǎ > o | + | + | + | + | + | + | + | + | + |
| * ě > o | – | + | + | – | – | + | – | – | – |
| АКАННЯ | – | + | + | – | – | – | – | – | – |
| el > olo | + | + | + | – | – | – | – | – | – |
| el > le | – | – | – | + | – | + | – | – | + |
| el > lie | – | – | – | – | + | – | – | + | – |
| ьг > og | + | + | + | – | – | – | – | – | – |
| ьг > ag | – | – | – | – | – | + | – | – | – |
| ль > l | + | + | + | + | + | + | + | + | + |
| ль > ly | + | + | + | + | + | + | – | – | – |
| ль > l'o | + | + | + | – | – | – | – | – | – |
| x > x' | – | + | – | – | – | – | – | – | + |
| x > š | + | + | + | + | + | + | + | + | + |

і т.д., усього 180 рядків.

4. Для кожної пари зіставляваних мов будують таблицю спряжених ознак 2×2 (див. Табл. 1). У ній чотири поля: **a**, яке характеризується двома плюсами, тобто, містить ДО, притаманні обом мовам, **b** (+ -), ознака властива першій мові й не властива другій, **c** (- +) ДО властива другій і не властива першій, **d** (- -) ДО не властива обом мовам.

Таким чином, ознаки, що містяться в полях **a** і **d**, є об'єднувальними, оскільки вказують на однакові характеристики зіставляваних мов, а ДО в полях **b** і **c** є роз'єднувальними. Співвідношення цих двох груп ДО визначатиме ступінь подібності, або генетичної спорідненості, зіставляваних мов. Чим більше ДО в полях **a** та **d**, тим подібніші мови.

5. Вираховують кількість ДО в кожному з чотирьох полів, щоб використати відповідні числа у формулі, за якою обраховують коефіцієнт подібності, або генетичної спорідненості, зіставляваних мов **Q**. Можна побудувати проміжну таблицю, в якій у верхньому рядку вказано поля **a**, **b**, **c**, **d**, а лівий крайній стовпчик указує пари зіставляваних мов (Табл. 3).

6. Для кожної пари мов вираховують показник генетичної спорідненості за формулою:

$$Q = \frac{ad - bc}{ad + bc}$$

Будують зіставну таблицю показників генетичної спорідненості зіставляваних мов (Табл. 4).

Опишемо хід нашого дослідження.

Список досліджуваних слов'янських мов складається з дев'яти мов: українська (Укр.), російська (Рос.), білоруська (Блр.), чеська (Чеш.), словацька (Слц.), польська (Плс.), болгарська (Блг.), сербохорватська (Срх.), словенська (Слв.).

За науковими джерелами [1; 2] відібрано 180 диференційних ознак, що відображають фонетичні зміни, які відбувалися в процесі становлення та розвитку слов'янських мов, наприклад:

*ě > і Укр. — *міра, місяць*

*ě > е Укр. — немає, Рос. — *мера, месяц*, Блр. — *мера, месяц*

*ě > а Укр., Рос. — немає, Блр. — *вяра, звяр*

*ě > о Рос. — *мед*

х > š Укр., Рос., Блр. — *муха, мушка*

Дія цих та інших ДО різна: є такі, що властиві всім зіставляваним мовам, інші — частині мов, а ще інші — виокремлюють якусь одну мову. За результатами аналізу функціонування ДО будують таблицю, у якій знаками + і - відображають наявність чи відсутність кожного ДО у мові (Табл. 2).

Аналіз таблиці 2 показує, що ДО функціонують по-різному: одні є об'єднувальними, оскільки кожна пара зіставляваних мов характеризується або двома плюсами, або двома мінусами, що свідчить про

генетичну спорідненість. 30,56% ДО властиві всім дев'яти мовам, але немає жодної ДО, яка була б не властива всім мовам, тобто не характеризується б жодним плюсом. І це закономірно: якщо ДО не має плюсів, то вона не притаманна слов'янським мовам взагалі, тому використовувати її для класифікації цих мов недоцільно.

17,78% ДО виокремлюють одну мову за рахунок присвоєння їй знака +, тоді як усі інші мають знак —, або тому, що вона характеризується знаком —, а всі інші — знаком +. Частіше інших виокремлюються чеська й болгарська мови. Це можна пояснити тим, що обидві вони зазнали значного впливу: чеська — німецької, а болгарська — турецької мови, що знизило ступінь генетичної спорідненості цих мов з мовами слов'янськими.

Взагалі за кількістю плюсів і мінусів, присвоєних аналізованим мовам, ДО розподілилися так:

| | | |
|--------------------|---------------------|---------------------|
| 9 плюсів, мінусів | — немає | — 55 ДО, або 30,56% |
| 8 плюсів, 1 мінус | — 11 ДО, або 6,11% | |
| 7 плюсів, 2 мінуси | — 6 ДО, або 3,33% | |
| 6 плюсів, 3 мінуси | — 13 ДО, або 7,22% | |
| 5 плюсів, 4 мінуси | — 17 ДО, або 9,44% | |
| 4 плюси, 5 мінусів | — 11 ДО, або 6,11% | |
| 3 плюси, 6 мінусів | — 25 ДО, або 13,89% | |
| 2 плюси, 7 мінусів | — 21 ДО, або 11,67% | |
| 1 плюс, 8 мінусів | — 21 ДО, або 11,67% | |
| Разом | — 180 ДО, або 100% | |

Взагалі в таблиці 2 є 1620 позицій (180 x 9), із них у 970 маємо плюс (59,88%) і в 650 — мінус (40,12%).

Отже, генетична спорідненість слов'янських мов більшою мірою зумовлена тим, що ДО властиві мовам, ніж тим, що вони їм не властиві.

Цікаво, що у випадку, коли виокремлюються три мови за генетичною спорідненістю, то частіше за все це східнослов'янські, які протиставляються іншим 12 разів, з них 11 — за знаком плюс. Західно- та південнослов'янські виокремлюються всього за чотирма ДО кожна. Це спостереження наводить на думку, що ступінь спорідненості східнослов'янських мов вищий, ніж мов двох інших підгруп.

Отже, як уже зазначалося раніше, основними для визначення ступеня генетичної спорідненості мов є показники полів **a** та **d**, які відображають однакову дію ДО в зіставлених мовах, тому чим більша їх кількість, тим більший ступінь спорідненості цих мов.

Визначення ступеня генетичної спорідненості аналізованих мов здійснюється шляхом побудови таблиць спряжених ознак, як про це йшлося вище. Щоб полегшити побудову таблиць 2×2, укладаємо проміжну узагальнену таблицю, в якій наводяться дані про кількість показників **a**, **b**, **c**, **d** для кожної пари мов. Вона не тільки полегшує побудову таблиць 2×2, а й забезпечує перевірку результатів майбутніх підрахунків за даними цих таблиць.

Ступені генетичної спорідненості української мови з іншими слов'янськими...

Таблиця 3. Кількість a, b, c, d у парах мов.

| Пари мов | a | b | c | d |
|----------|-----|----|----|----|
| Укр-Рос | 101 | 17 | 16 | 46 |
| Укр-Блр | 109 | 10 | 10 | 51 |
| Укр-Чеш | 78 | 39 | 27 | 36 |
| Укр-Слц | 77 | 41 | 23 | 39 |
| Укр-Плс | 80 | 38 | 24 | 38 |
| Укр-Блг | 81 | 37 | 21 | 41 |
| Укр-Срх | 81 | 37 | 18 | 44 |
| Укр-Слв | 74 | 43 | 22 | 41 |
| Рос-Блр | 107 | 10 | 13 | 50 |
| Рос-Чеш | 76 | 41 | 29 | 34 |
| Рос-Слц | 80 | 38 | 22 | 40 |
| Рос-Плс | 82 | 35 | 23 | 40 |
| Рос-Блг | 82 | 35 | 21 | 42 |
| Рос-Срх | 79 | 38 | 21 | 42 |
| Рос-Слв | 76 | 41 | 21 | 42 |
| Блр-Чеш | 76 | 43 | 29 | 32 |
| Блр-Слц | 80 | 39 | 21 | 40 |
| Блр-Плс | 84 | 35 | 21 | 40 |
| Блр-Блг | 78 | 41 | 23 | 38 |
| Блр-Срх | 79 | 40 | 20 | 41 |
| Блр-Слв | 72 | 47 | 25 | 36 |
| Чеш-Слц | 90 | 15 | 12 | 63 |
| Чеш-Плс | 84 | 21 | 21 | 54 |
| Чеш-Блг | 78 | 27 | 24 | 51 |
| Чеш-Срх | 77 | 28 | 22 | 53 |
| Чеш-Слв | 76 | 28 | 20 | 56 |
| Слц-Плс | 81 | 20 | 24 | 55 |
| Слц-Блг | 76 | 25 | 26 | 53 |
| Слц-Срх | 79 | 22 | 20 | 59 |
| Слц-Слв | 76 | 25 | 21 | 58 |
| Плс-Блг | 74 | 31 | 28 | 47 |
| Плс-Срх | 69 | 36 | 29 | 46 |
| Плс-Слв | 69 | 36 | 28 | 47 |
| Блг-Срх | 81 | 21 | 18 | 60 |
| Блг-Слв | 77 | 25 | 19 | 59 |
| Срх-Слв | 85 | 14 | 11 | 70 |

Аналіз цієї таблиці показує, що найвищі показники **a** та **d** припадають на пари східнослов'янських мов: Укр-Рос 147, Укр-Блр 160, Рос-Блр 157. Цього рівня досягає лише пара Чеш-Слц 153. Отже, ці дані підкріплюють нашу думку про найвищий ступінь генетичної спорідненості східнослов'янських мов.

Наступний етап — побудова таблиць 2×2 для всіх 36 пар аналізованих мов. Наведемо два приклади таких таблиць: для мов з найвищим і найнижчим показником **a**.

| | | Укр-Блр | | |
|-----------|--|---------|----|-------|
| Укр \ Блр | | + | - | Разом |
| + | | 109 | 10 | 119 |
| - | | 10 | 51 | 61 |
| Разом | | 119 | 61 | 180 |

| | | Плс-Срх | | |
|-----------|--|---------|----|-------|
| Плс \ Срх | | + | - | Разом |
| + | | 69 | 36 | 105 |
| - | | 29 | 46 | 75 |
| Разом | | 98 | 82 | 180 |

Виразуємо коефіцієнт генетичної спорідненості Q для кожної з цих пар:

$$Q_{\text{Укр-Блр}} = \frac{109 \times 51 - 10 \times 10}{109 \times 51 + 10 \times 10} = \frac{5559 - 100}{5559 + 100} = \frac{5459}{5659} = 0,97$$

$$Q_{\text{Плс-Срх}} = \frac{69 \times 46 - 29 \times 36}{69 \times 46 + 29 \times 36} = \frac{3174 - 1044}{3174 + 1044} = \frac{2130}{4218} = 0,51$$

Як бачимо, дійсно, величина коефіцієнта спорідненості мов зменшується зі зменшенням показника a .

Звертає на себе увагу той факт, що в таблицях 2×2 наявні суми $a+b$, $a+c$, $b+d$, $c+d$, які не використовуються в нашій формулі. Справа в тому, що існує інша формула для вираховування Q , де ці суми використовуються:

$$Q = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$$

Виразуємо Q для мов Укр-Блр і Плс-Срх за цією формулою:

$$Q_1 = \frac{5559 - 100}{\sqrt{119 \times 119 \times 61 \times 61}} = \frac{5459}{\sqrt{52693081}} = \frac{5459}{7263} = 0,75$$

$$Q_2 = \frac{2130}{\sqrt{105 \times 98 \times 75 \times 82}} = \frac{2130}{7912} = 0,27$$

Як бачимо, числові значення Q менші, але це не означає, що змінився ступінь спорідненості, просто, обчислення тут ведеться, так би мовити, в іншому масштабі. І це треба враховувати при зіставленні коефіцієнтів: зіставляти можна лише ті, які вираховані за тією самою формулою. Ми будемо використовувати першу формулу, за якою обраховано всі коефіцієнти, наведені в Табл. 4.

Найвищі коефіцієнти спорідненості мають східнослов'янські мови. Показник Q для східнослов'янських мов знаходиться в межах 0,97 – 0,89, тоді як для західнослов'янських його межі становлять 0,93 – 0,80, а для південнослов'янських 0,94 – 0,81. Отже, південнослов'янські мови ближчі до східнослов'янських, ніж західнослов'янські.

Ступені генетичної спорідненості української мови з іншими слов'янськими...

Таблиця 4. Коефіцієнти генетичної спорідненості слов'янських мов.

| | Укр. | Рос. | Блр. | Чеш. | Слц. | Плс. | Блг. | Срх. | Слв. |
|------|------|------|------|------|------|------|------|------|------|
| Укр. | | 0,89 | 0,97 | 0,45 | 0,50 | 0,54 | 0,62 | 0,69 | 0,52 |
| Рос. | | | 0,95 | 0,37 | 0,59 | 0,60 | 0,65 | 0,61 | 0,56 |
| Блр. | | | | 0,32 | 0,59 | 0,64 | 0,52 | 0,60 | 0,37 |
| Чеш. | | | | | 0,93 | 0,82 | 0,72 | 0,73 | 0,77 |
| Слц. | | | | | | 0,80 | 0,72 | 0,83 | 0,78 |
| Плс. | | | | | | | 0,60 | 0,51 | 0,52 |
| Блг. | | | | | | | | 0,85 | 0,81 |
| Срх. | | | | | | | | | 0,94 |
| Слв. | | | | | | | | | |

Західнослов'янські мови теж об'єднуються за спорідненістю, але на такому ж рівні спорідненість словацької мови із сербохорватською. Їх коефіцієнти спорідненості дещо нижчі, лише чеська та словацька мови мають Q вищий за 0,90.

Таку ж картину спостерігаємо й у південнослов'янських мовах: лише одна пара мов, сербохорватська і словенська, мають високий коефіцієнт спорідненості.

Спорідненість західнослов'янських мов зі східнослов'янськими описується Q , що знаходиться в межах 0,64 — 0,32, а для південнослов'янських Q становить від 0,69 до 0,37. Отже, цей показник свідчить, що за генетичною спорідненістю східнослов'янські мови становлять ядро слов'янських мов, від якого південно- та західнослов'янські мови перебувають на різних відстанях.

По відношенню до західнослов'янських мов південнослов'янські мають показник Q від 0,83 до 0,51.

Отже, ієрархію слов'янських мов можна зобразити так:



Найнижчі коефіцієнти спорідненості у чеської мови з російською та білоруською мовами, що можна пояснити значним впливом німецької мови на чеську; так само низька спорідненість характерна для білоруської та словенської мов. Взагалі спорідненість чеської мови зі східнослов'янськими мовами нижча, ніж з південнослов'янськими.

Розподіл мов за ступенем генетичної спорідненості має таку картину:

| | |
|------------------------|---------------------------------|
| 0,97– Укр-Блр | 0,65– Рос-Блг |
| 0,95– Рос-Блр | 0,64– Блг-Плс |
| 0,94– Срх-Слв | 0,62– Укр-Блг |
| 0,93– Чеш-Слц | 0,61– Рос-Срх |
| 0,89– Укр-Рос | 0,60– Рос-Плс, Плс-Блг, Блр-Срх |
| 0,85– Блг-Срх | 0,59– Рос-Слц, Блр-Слц |
| 0,83– Слц-Срх | 0,58– Рос-Слв |
| 0,82– Чеш-Плс | 0,54– Укр-Плс |
| 0,81– Блг-Слв | 0,52– Укр-Слв, Блр-Блг, Плс-Слв |
| 0,80– Слц-Плс | 0,51– Плс-Срх |
| 0,78– Слц-Слв | 0,50– Укр-Слц |
| 0,77– Чеш-Слв | 0,45– Укр-Чеш |
| 0,73– Чеш-Срх | 0,37– Рос-Чеш, Блр-Слв |
| 0,72– Чеш-Блг, Слц-Блг | 0,32– Блр-Чеш |
| 0,69– Укр-Срх | |

Цікаво, що всього п'ять угруповань за ступенем генетичної спорідненості: кожна пара мов характеризується своїм коефіцієнтом.

У порівняльно-історичному мовознавстві генетична спорідненість базується на аналізі фонетичних, граматичних і лексичних змін протягом історичного розвитку мови. У цьому дослідженні враховано тільки фонетичні зміни. Як уявляється, корисно провести дослідження окремо для кожної групи ДО і з'ясувати, які ДО і як саме впливають на ступінь генетичної спорідненості мов. Якщо виявиться, що їх дія збігається, об'єднання результатів цих досліджень, можливо, уточнять наш результат.

1. Вступ до порівняльно-історичного вивчення слов'янських мов // За ред. О.С. Мельничука. — К.: Наук. думка, 1966. — 588 с.
2. *Лекс И.* Насоки в развоја на фонолоичните системи на славјанските езици. — Софија, 1960.
3. *Перебийніс В.І.* Статистичні методи для лінгвістів. — Вінниця: Нова книга, 2002. — 168 с.

Valentina Perebeynoss (Kyiv)

DEGREES OF GENETIC AFFINITY OF THE UKRAINIAN LANGUAGE
FOR OTHER SLAVIC LANGUAGES: A STATISTICAL APPROACH.

A statistical approach to establishing degrees of genetic affinity of Slavic languages is described. It is based on a comparison of languages in pairs by differential features (DF) reflecting phonetical laws of their historical development. If a DF is characteristic of both compared languages it is called a uniting DF. The DF characteristic only of one of the two languages is called a separating DF. The more uniting differential features are peculiar to the compared languages, the higher the degree of their genetic affinity is.

Key words: Slavic languages, genetic affinity, a uniting differential feature, a separating differential feature, statistical approach, a table of conjugate features.