А.А. Соловьева

ПРОФЕССИОНАЛЬНАЯ НАПРАВЛЕННОСТЬ В ОБУЧЕНИИ МАТЕМАТИКЕ СТУДЕНТОВ СПЕЦИАЛЬНОСТИ «ЛИНГВИСТИКА»

…имеющие уважение к слову должны знать,
что существует математика слова
А. Блок

Привлечение математического аппарата в языкознании началось еще в начале прошлого столетия. Тогда проводились первые попытки статистической обработки текста [5], [6], [7] и др., которые заключались в подсчете словоупотреблений в произведениях разных авторов или в текстах одного автора и сравнении полученных частот. Задачами подобных исследований являлись определение авторства или выявление плагиата. Бодуэн де Куртенэ в 1904 году писал: «Поскольку в языкознании применяются количественные понятия, желательно тоже знание математики, не только низшей, но и высшей…».

Сегодня для развития теории и практики языкознание привлекает методы из значительного большинства разделов математики, а области использования этих методов расширяются. Математические методы применяются для исследований изменений языковых процессов во времени (диахроническая лингвистика), географическом пространстве (диалектология), специально-профессиональном и художественном континууме (социолингвистика и стилистика), для решения прикладных вопросов, связанных с анализом, синтезом и информационной обработкой текста (компьютерная лингвистика) и др. Среди компьютерных программ особое место занимают лингвистические программы (*Link Grammar Parser*, *Лингвоанализатор, TextAnalyst* 2.0, *Concordance* 2.0.0 и др. [2]), как специализированные для проведения конкретных исследований, так и созданные для широкого круга пользователей.

В данной статье предлагается идея использования спиралей фундирования базовых учебных элементов в процессе преподавания математики студентам-лингвистам. В ее основе - усиление профессионального компонента математического образования с последующим фундированием знаний на разных уровнях. В результате реализуются две задачи обучения: 1) обеспечение тесных связей математических задач с задачами профессионального развития будущего специалиста; 2) поддержание высокого уровня мотивации. Эта идея впервые была представлена в книге «Подготовка учителя математики: инновационные подходы» [10] и продолжена в [1], [11], [12] на процесс преподавания математических дисциплин студентам гуманитарных специальностей.

Развертывание спирали фундирования базовых учебных элементов при преподавании математики на специальности «лингвистика» можно разбить на три этапа. На *первом* рассматриваются представления о понятии, используемом в языкознании, которые иногда бывают интуитивными. На *втором этапе* осуществляется процесс фундирования этого понятия в математической дисциплине. Данный процесс реализуется с помощью нескольких шагов, на каждом из которых рассматриваются понятия и утверждения, необходимые для введения конечного понятия. На *третьем этапе* происходит проецирование теоретического материала на будущую деятельность студента в форме актуализированных практических приложений. Таким образом, в основе спиралевидной схемы развертывания и моделирования базовых учебных элементов лежит выделение профессионально-ориентированного теоретического материала.

Реализацию данной схемы рассмотрим на примерах спиралей фундирования двух базовых учебных элементов: понятия функции в лингвистике (рис. 1) и понятия предела функции в лингвистике (рис. 2).

Первый этап. Текст можно рассматривать как линейную цепочку отграниченных друг от друга символов (фонем, букв, слогов, слов). Символы образуют определенные совокупности. Каждый из символов встречается в тексте с определенной частотой и обладает особыми «способностями сочетаться с другими символами» [8. С.9]. Также существуют и такие лингвистические явления, как употребительность слова или словосочетания, длина звука, длина буквосочетания, информационный вес слога, морфемы или слова, степень аналитичности языка. Подобные явления можно выражать в виде чисел, а значит, и рассматривать в качестве математических величин.

Спираль фундирования понятия функции в лингвистике

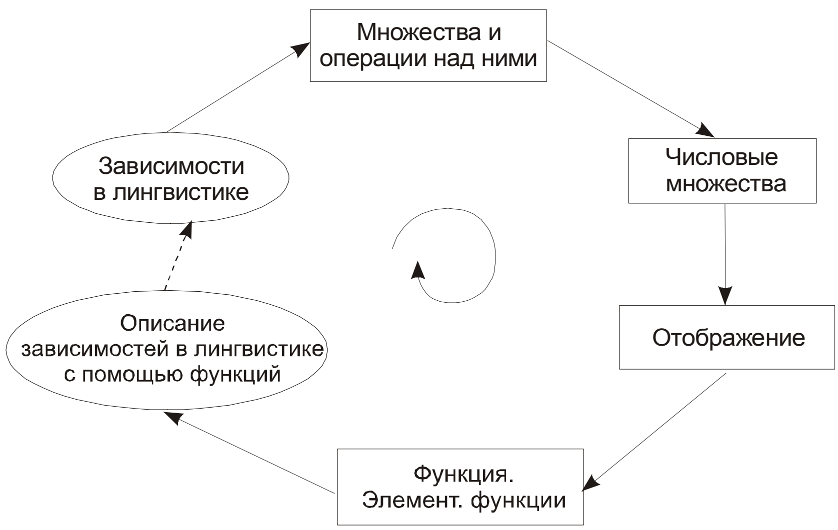

Рис. 1

Между различными лингвистическими явлениями наблюдаются определенные зависимости. Примерами являются зависимость между индоевропейскими звонкими смычными придыхательными согласными и их готскими производными [8. С. 22]; зависимость между частотой словоформы, которую она имеет в тексте определенной длины, и ее номером в частотном словаре; употребительность лингвистического явления на различных этапах его истории [8. С. 31-38]. Для указанных и других зависимостей существует возможность их описания средствами математического аппарата, а они, в свою очередь, открывают новые методы для решения задач лингвистики.

Второй этап. Для математического описания зависимостей между лингвистическими явлениями используются понятия отображения и функции, этап фундирования в математике которых начинается с рассмотрения *понятия множества*. В процессе введения этого понятия даются определения пустого, конечного, бесконечного, несобственного и собственного множеств, эквивалентности множеств, подмножества множества, рассматривается, что значит задать множество, и способы его задания.

Введение понятий дополняется примерами из лингвистики. Примерами множеств могут быть множество букв русского алфавита, множество словоупотреблений, содержащихся на данной странице. Пустым множеством является множество двухбуквенных комбинаций *сй*, *уь*, *чя*, если иметь в виду русские тексты, не содержащие опечаток. Конечные множества в лингвистике встречаются чаще всего, хотя приходится рассматривать и бесконечные, примером которых является «множество всех словоупотреблений в текстах данного языка при условии, что этот язык беспрерывно порождает и будет порождать новые тексты без ограничения во времени» [8. С.13]. Множество губно-зубных твердых и мягких согласных {[ф], [ф'], [в], [в']} можно считать подмножеством множества губных согласных {[п], [п'], [б], [б'], [м], [м'], [ф], [ф'], [в], [в']} [3]. Сравнивая множество *A*, состоящее из словоформ *они*, *их*, *им*, *ими*, со множеством *B*, содержащим формы склонения местоимения *они*, убеждаемся, что множества *A* и *B* – эквивалентные.

Далее рассматриваются *основные операции над множествами* (объединение, пересечение, разность) и их свойства. В качестве примера объединения множеств можно привести следующий.

Пример 1. Рассмотрим два множества: множество сонорных согласных, $A =$ {[м], [м'], [н], [н'], [л], [л'], [р], [р']} – согласных, образуемых при помощи голоса и незначительного шума, и множество губно-губных согласных $B =$ { [б], [б'], [п], [п'], [м], [м']}. Объединением этих множеств является множество $A \cup B =$ {[б], [б'], [п], [п'], [м], [м'], [н], [н'], [л], [л'], [р], [р']}, пересечением множеств – множество $A \cap B =$ {[м], [м']}, разностью множеств *B* и *A* – множество $B \setminus A =$ { [б], [б'], [п], [п']}.

Наряду с понятием множества квантитативная лингвистика, исследующая количественную сторону языка и речи, постоянно оперирует и понятием числа. *Числовые множества* находят свое применение для описания лингвистических явлений и процессов. Натуральные числа: в каждом слове имеется целое положительное число букв, в предложении – целое положительное число слов и т.п. Однако натурального ряда чисел недостаточно, например, для измерения средней встречаемости той или иной грамматической, лексической, фонологической единицы используются дробные числа. В некоторых лингвистических зависимостях используются и отрицательные величины. Квантитативные измерения текста не ограничиваются четырьмя действиями (сложение, вычитание, умножение, деление) элементарной математики, в лингвистике приходится решать задачи, требующие использования, например, логарифмирования (исследования информационного веса лингвистических единиц). Запаса множества действительных чисел (рациональные и иррациональные) достаточно для решения основных задач квантитативной лингвистики.

Следующий шаг второго этапа – введение *понятия отображения*. Здесь рассматривается, что значит задать отображение, понятия взаимно однозначного (биективного), инъективного, сурьективного отображения, оператора. В качестве примера отображения в лингвистике можно рассмотреть соотношение индоевропейских и готских согласных.

Пример 2. Индоевропейские звонкие смычные придыхательные согласные могут быть традиционно [4. С.11-116] представлены в виде множества $A = \left\{ bh,\ dh,\ gh \right\}$, а соответствующие им готские согласные объединены во множество $B = \left\{ b,\ d,\ g \right\}$. Каждому звукотипу во множестве *A* однозначно соответствует определенный звукотип во множестве *B* по определенному правилу: каждый из индоевропейских звукотипов теряет придыхательность. Значит, можно сказать, что задано отображение $f : A \to B$ из множества *A* во множество *B*, которое является еще и взаимно однозначным.

После рассмотрения понятия отображения вводится *понятие функции*, рассматриваются способы задания функции. Примером функции в лингвистике является формула

$$p(x) = \frac{k}{(x+\rho)^{\gamma}},$$

которая описывает соответствие между множеством номеров словоформ частотного списка и множеством их вероятностей $p(x)$. Величина x – независимая переменная, $p(x)$ – функция, величины k, ρ, γ – постоянные из множества действительных чисел [8. С.24].

Рассмотрим еще один пример, в котором указано обобщенное описание особенностей пушкинских ямбов.

Пример 3. Б.В. Томашевским было проведено статистическое исследование достаточно большого числа пушкинских текстов. В результате было установлено, что между числом стоп (четных слогов) в ямбе (x) и средним количеством пиррихиев в одной стихотворной строке (y) существует вполне определенное соответствие:

$$y = 0{,}28(x-1).$$

Это значит, что количество пиррихиев прямо пропорционально числу четных слогов в строке минус слог рифмующий, так как он не участвует в распределении пиррихиев.

В заключение второго этапа фундирования понятия функции в лингвистике рассматриваются основные *элементарные функции*: линейная, квадратичная, степенная, дробно-рациональная, показательная, логарифмическая, тригонометрические, обратные тригонометрические и их графики.

Третий этап. Описание многих лингвистических явлений с помощью понятия функции расширяет диапазон методов для проведения лингвистических исследований.

Пример 4. Побуквенное распределение синтактической информации I в слове можно представить в виде следующего аналитического выражения:

$$I(n) = I_0 e^{-sn},$$

где в качестве аргумента выступает номер n буквы слова, $I(n)$ - количество информации, которое несет буква n, параметр I_0 (называемый информацией алфавита) показывает максимальную величину информации, которую несла бы буква языка, использующего алфавит S, параметр s показывает темп нарастания ограничений, накладываемых системой и нормой языка на неопределенность выбора n-й буквы слова при условии, что цепочка букв, находящихся слева, уже известна [9. С.80-89]. Смысл данного выражения в том, что начальные буквы письменного слова несут значительно больше информации, чем буквы, находящиеся в его середине и на конце.

Другие элементарные функции также используются для описания лингвистических явлений: с помощью показательной функции строятся модели распределения информации в слове, в тексте, контекстной обусловленности; логарифмическая функция используется при построении математического описания количества информации словаря [9]; сумма тригонометрических функций $y = \sin\ x$ с введенными определенными коэффициентами отражает структуру гласных звуков [8. С.41-47]; с помощью обратных тригонометрических функций моделируется диахронический скачок и т.п.

Рассмотрим пример моделирования развития нулевых форм родительного падежа множественного числа у русских единиц измерения типа *вольт*, *рентген*, *радиан* [8. С.32-34].

Пример 5. Процесс формирования этой новой лексико-грамматической группы имен существительных, которая имеет в родительном падеже множественного числа нулевое окончание, развивался следующим образом. В русских научно-технических текстах XIX в. употреблялись регулярные образования родительного падежа множественного числа: *вольтов*, *рентгенов*. Начиная с конца 80-х годов того же века, отмечалось нарастание употребления форм: *вольт*(ъ), *рентген*(ъ), *радиан*(ъ), совпадающих с именительным падежом единственного числа. Через 20-30 лет новые формы утвердились не только в профессиональной речи, но и в литературном языке.

Статистический ход этого процесса по годам (t) представлен в табл. 1, где p – относительные частоты нулевых форм.

Таблица 1

t	1881	1885	1887	1889	1891	1895	1897	1899	1907	1910	1901
p	0,06	0,01	0,03	0,11	0,57	0,47	0,68	0,96	0,92	0,99	0,95

Для анализа процесса следует построить его аналитическую модель $p = f(t)$. Выясним, какой из известных элементарных функций можно воспользоваться для описания полученной зависимости.

Если провести эмпирическую кривую, то становится видно, что значения аргумента t теоретически могут располагаться по всей числовой прямой, а значения функции лежат в интервале [0, 1]. Левая ветвь асимптотически приближается к оси Ot, а правая – к прямой $p = 1$. В середине кривой в промежутке между 1886 и 1905 годами график резко возрастает. Рассмотренным условиям больше всего соответствует график функции $y = \operatorname{arctg}\ x$. Учитывая с помощью дополнительных коэффициентов все особенности кривой, получаем следующее аналитическое выражение:

$$p(t) = \frac{1}{\pi} \operatorname{arctg} \left(\frac{t - 1895}{3} \right) + 0{,}5\,.$$

Построенную модель можно использовать при описании структурных сдвигов в области лексики, морфологии, фонологии, синтаксисе, стилистике. Такие сдвиги обнаруживаются либо в появлении новых элементов, либо в исчезновении старых. Так как структурные изменения языка при этом происходят скачкообразно, подобные явления в лингвистике носят название диахронического скачка.

Моделирование таким образом развития лингвистического процесса с помощью элементарных функций позволяет решать некоторые теоретические вопросы языкознания, например, служит средством восстановления не засвидетельствованных в памятниках и диалектах этапов исследуемого процесса.

Модель процесса формирования определенного артикля во французском языке (V-XIII в.) [8.С. 34-38]

$$p(t) = \frac{0{,}4}{\pi} \operatorname{arctg} \left(\frac{t - 1125}{100} \right) + 0{,}2$$

помогла при описании некоторых деталей начального периода развития исследуемого явления, что невозможно было сделать непосредственно по народнолатинским текстам V-VIII в.

После введения понятия функции в математике и знакомства с его применениями в лингвистике предлагается рассмотреть понятие предела функции, спираль фундирования которого представлена на рис. 2.

Спираль фундирования понятия предела в лингвистике

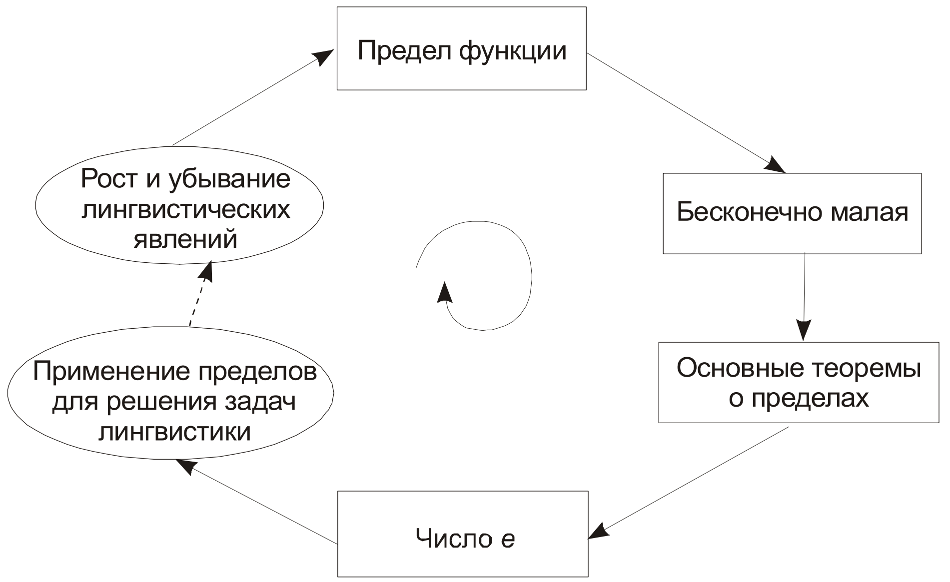

Рис. 2

Первый этап. Среди лингвистических процессов существуют такие, которые являются непрерывно текущими (мало изменяющимися в малый момент времени) и неоднородными по своему темпу. К ним относятся процессы роста словаря, распада языка, построения информационной схемы слова, связного текста и др. Подобные диахронические и информационные процессы характеризуются ростом или убыванием какой-либо величины.

Для описания и анализа таких процессов следует применять более гибкий по сравнению с предыдущим математический аппарат, построенный на понятиях предела и бесконечно малой величины.

Второй этап начинается с введения самого *понятия предела* в математике. Дается определение предела функции в точке, рассматриваются пределы функции при $x \to x_0$, $x \to \pm\infty$, вводится понятие предела последовательности.

Понятие предела можно проиллюстрировать на примере роста употребительности нулевых форм родительного падежа множественного числа у существительных, обозначающих единицы измерения: *вольт*, *рентген*, *радиан*.

Пример 6. В предыдущем примере для указанного процесса была получена аналитическая модель:

$$p(t) = \frac{1}{\pi} \operatorname{arctg} \left(\frac{t-1895}{3} \right) + 0{,}5 .$$

При t = 1905, 1907, 1908, …, 1920, …, 1950, функция $p(t)$ принимает соответствующие значения 0,906, 0,922, 0, 927, …, 0,962, …, 0,983. Видно, что по мере приближения t к ∞ рассматриваемая функция стремится к единице, т.е

$$\lim_{t \to \infty} p(t) = 1 .$$

Далее вводится *понятие бесконечно малой* функции и рассматриваются свойства бесконечно малых. Примером бесконечно малой может быть функция $p(n)$ вероятности того, что фонологическая система примет устойчивое положение при однократном случайном изменении одного из дифференциальных признаков этой системы [8].

Пример 7. Указанная вероятность определяется выражением

$$p(n) = \frac{1}{2^n} .$$

При увеличении числа дифференциальных признаков фонологической системы (n = 0, 1, 2, 3, …) вероятность принимает соответствующие значения 1, $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{16}$, … Какое бы малое положительное число ε мы ни взяли, найдется такое *n*, при котором значений *p* будет меньшее, чем ε. Пусть $\varepsilon = \frac{1}{1000}$, то при $n = 10$ *p* принимает значение $\frac{1}{1024}$, меньшее, чем ε. При $\varepsilon = \frac{1}{10000}$ находится $n = 14$, при котором $p = \frac{1}{16384}$ снова меньше ε. Таким образом, функция $p(n) = \frac{1}{2^n}$ - бесконечно малая.

Следующим шагом является рассмотрение *основных теорем о пределах*: единственность предела; теорема о пределе промежуточной функции; теорема о представлении функции в виде суммы числа, равного пределу этой функции в некоторой точке, и бесконечно малой в этой же точке; теорема об операциях над пределами.

В заключение рассматривается предел $\lim\limits_{x\to\infty}\left(1+\frac{1}{x}\right)^x$, равный числу Эйлера *e* = 2,718…,и пример применения числа *e* для построения идеальной модели роста словаря [8.С. 56-57].

Пример 8. В результате постоянного расширения сферы деятельности человека его лексический запас неуклонно растет. Характеристикой увеличения словаря служит коэффициент k его прироста за определенный период времени:

$$k = \frac{\Delta L}{L},$$

где ΔL – количество новых слов, появившихся за данный период времени, за вычетом вышедших из употребления архаизмов, L – общий объем словаря на данный период времени. Зная начальный объем словаря L и коэффициент *k*, легко показать, что через 10 лет объем словаря составит

$$L_0 + \Delta L = L_0 + L_0 k = L_0(1+k).$$

Однако такой подсчет дает грубое приближение, так как в течение десятилетия прирост словаря происходит не относительно исходной величины L_0, а относительно сумм $L_0 + \Delta L$, где ΔL указывает на прирост словаря за год, месяц, неделю и т.д. Учитывая рост словаря по годам, выведем более точный результат. К концу первого года получаем

$$L_1 = L_0\left(1+\frac{k}{10}\right),$$

к концу второго года имеем

$$L_2 = L_1 + L_1\frac{k}{10} = L_0\left(1+\frac{k}{10}\right)^2,$$

а к концу десятилетия объем словаря составит

$$L_{10} = L_0\left(1+\frac{k}{10}\right)^{10}.$$

При учете прироста словаря по месяцам результат объема словаря к концу десятилетия оказывается еще более точным

$$L_{10} = L_0\left(1+\frac{k}{120}\right)^{120}.$$

Теперь обобщим процесс роста словаря. Промежуток времени T разделим на *n* равных частей. Объем словаря к концу периода T будет составлять

$$L_T = L_0\left(1+\frac{kT}{n}\right)^n.$$

Введем условие неограниченного возрастания числа промежутков *n*, получаем

$$L_T = \lim_{n\to\infty} L_0\left(1+\frac{kT}{n}\right)^n = L_0 \lim_{n\to\infty}\left[\left(1+\frac{kT}{n}\right)^{\frac{n}{kT}}\right]^{kT} = L_0\left[\lim_{n\to\infty}\left(1+\frac{kT}{n}\right)^{\frac{n}{kT}}\right]^{kT}$$

Получаем идеальную модель роста словаря, выраженную формулой

$$L_T = L_0 e^{kT}.$$

Третий этап. С помощью понятия предела решаются некоторые задачи квантитативной лингвистики, связанные с ростом или убыванием какой-либо величины в диахронических или информационных процессах.

Например, процесс распада некоторого языка характеризуется убыванием числа слов, обозначающих необходимые понятия (например, *маленький*, *живот-*

ное, *все* и др.), и представляет собой непрерывно текущий (мало изменяющийся в малый момент времени) лингвистический процесс. С целью определения приближенной датировки расхождения диалектов и родственных языков, а также для количественной оценки степени их родства применяется описанный математический аппарат [8.С. 57-59]. Подобные исследования в своей основе имеют пять утверждений, которые обозначены как постулаты:

1. Во всех языках мира существует некоторое множество слов, обозначающих наиболее древние, всегда необходимые и поэтому не изменяющиеся понятия-означаемые, называемое текстовым списком (ТС).
2. Доля означающих (слов или эквивалентных словам устойчивых словосочетаний) постоянна в течение некоторого промежутка времени T и не зависит от способа выбора этих слов из ТС.
3. Каждый язык и диалект имеет свой коэффициент сохранности r относительно периода в 1000 лет. Величины r колеблются относительно периода в тысячу лет от 0,75 (для быстро развивающихся языков) до 0,91 (для «стабильных» языков), средняя величина $\bar{r}$ равна 0,81.
4. Все означающие из ТС данного языка имеют одинаковые шансы сохраниться на протяжении периода времени Т.
5. Вероятность означающего из ТС праязыка сохраниться в ТС первого потомка не зависит от вероятности сохраниться в ТС второго потомка.

Опираясь на рассуждения, приведенные в примере 8, получаем формулу для оценки числа L_T общих означающих из ТС праязыка, которое сохраняется в двух языках-потомках за период их независимого развития:

$$L_T = L_0 e^{kT}, \qquad (*)$$

где L_0 – число означающих ТС, которые характеризуют праязык в момент выделения из него языков-потомков i и j, k – коэффициент потери общих слов за некоторый период T (называемый периодом дивергенции). Этот коэффициент равен $k = 1 - r_i r_j$, где r_i и r_j – коэффициенты сохранности лексики каждого из сравниваемых языков.

Прологарифмировав выражение (*), получаем

$$\ln L_T = \ln L_0 - kT,$$

значит, количественная оценка периода дивергенции определяется из равенства

$$T = \frac{\ln L_0 - \ln L_T}{k}. \qquad (**)$$

Пример 9. Известно, что из L_0 = 202 латинских слова дакорумынский и арумынский языки сохраняют только L_T = 149 общих лексем, а коэффициенты сохранности лексики соответственно r_d = 0,81 и r_a = 0,88. Определим период дивергенции для данной пары романских языков.

Коэффициент потери общих слов в ходе дивергенции равен

$$k = 1 - r_d r_a = 1 - 0{,}81 \cdot 0{,}88 \approx 0{,}29.$$

Тогда период дивергенции получаем из равенства (**):

$$T = \frac{\ln 202 - \ln 149}{0{,}29} = 1{,}05 \text{ тысячелетия.}$$

Таким образом, момент распада балканороманского языка относится к IX в.

Степень родства языков и диалектов оценивается через отношение $\gamma = \frac{L_T}{L_0}$. При γ≥0,81 два языка следует считать диалектами, при 0,36≤ γ <0,81 языки являются родственными, при 0,12≤ γ <0,36 принадлежат к одной ветви. Так, языки из примера 9 являются родственными.

Введение (изучение?) базовых учебных элементов курса математики студентам специальности «лингвистика», построенное по спиралевидным схемам, описанным выше, создает позитивную познавательную и профессиональную основу для будущей деятельности студента. Это обусловлено тем, что, с одной стороны, позволяет дать ему достаточную математическую подготовку, а с другой – навыки использования полученных знаний для решения конкретной задачи теории и практики.

Библиографический список

1. Афанасьев В.В., Соловьева А.А. Спираль фундирования понятия информации при преподавании математики студентам специальности «реклама». Сб. материалов научн.-метод. конференции «Чтения Ушинского». Ярославль, 2003.
2. Каталог лингвистических программ и ресурсов в Сети. www.rvb.ru/soft/catalogue/catalogue.html
3. Классификация русских согласных по месту образования. www.pholol.msu.ru/rus/lena-1/conson/mesto1.html
4. Мейе А. Сравнительное изучение индоевропейских языков. М. – Л.. Сосэкгиз, 1938.
5. Марков А.А. Об одном применении статистического метода // Известия Императорской Академии Наук, 1916.
6. Марков А.А. Применение статистическаго исследования // Известия Императорской Академии Наук, 1913.
7. Морозов Н.А. Лингвистические спектры: средство для отличения плагиатов от истинных произведений того или иного неизвестного автора. Стилеметрический этюд // Известия отд. русского языка и словесности Имп.Акад.наук. Т.XX. Кн.4. 1915.
8. Пиотровский Р.Г. и др. Математическая лингвистика: Учеб. пособие для пед. ин-тов. М., 1977.
9. Пиотровский Р.Г. Информационные измерения языка. Л.: Наука, 1968.
10. Подготовка учителя математики: инновационные подходы: Учеб. пособие / Под ред. В.Д. Шадрикова. М.: Гардарики, 2002.
11. Соловьева А.А. Спираль фундирования «Риск принятия решений» при преподавании математики студентам специальности «менеджмент организации». Сб. материалов школы-семинара, посв. 100-летию А.Н. Колмогорова. Ярославль, 2003.
12. Afanas’ev V.V., Solov’eva A.A. Using spirals of foundiration in the process of teaching mathematics to students of humanitarian specialities. 2003.