

СТЕПЕНЬ ВЗАИМНОЙ СИСТЕМНО-СИНЕРГЕТИЧЕСКОЙ БЛИЗОСТИ ЧЕТЫРЕХ РОМАНСКИХ ЯЗЫКОВ

*Работа представлена кафедрой романской филологии.
Научный руководитель – доктор филологических наук, профессор Р. Г. Пиотровский*

Статья посвящена определению степени системно-синергетической близости испанского, французского, итальянского и румынского языков на основании лексико-статистических данных. Результаты работы можно использовать для внутрилингвистической таксономии романских языков, которая отличается от традиционных классификаций, построенных на внешних критериях или частных лингвистических индикаторах.

Ключевые слова: романские языки, системно-синергетическая близость, лексико-статистические данные, частотность, таксономия, классификация.

The aim of the article consists in determining the degree of the system-synergic proximity among Spanish, French, Italian and Rumanian on the basis of the lexicostatistical data. The results of this research can be used for the immanent (in sense of L. Hjelmslev) linguistic taxonomy of Romance languages, which differs from traditional classifications that use also extra-linguistic criteria.

Key words: Romance languages, system-synergic proximity, lexicostatistical data, frequency, taxonomy, classification.

Введение. В наших предыдущих работах¹ рассматривалась степень близости испанского, французского и итальянского языков. Здесь с применением тех же критериев рассматривается также соотношение этих языков с румынским. Не повторяя подробного описания процедуры исследования, напомним его основные принципы.

Работа проводилась в синхронии. Это означает, что в ней учитываются формальные и семантические сходства лексических единиц (ЛЕ) лишь на современном этапе их существования, вне прямой связи с условиями их происхождения. Вместе с тем при определенных затруднениях в данном исследовании использовался этимологический словарь романских языков.

Методика исследования выглядит следующим образом. Для каждой пары языков (L_i , L_j) были построены сопоставительные таблицы. Одни из этих таблиц содержат номера ЛЕ, указанные в частотных словарях (ЧС), другие – те номера, которые получаются в каждой таблице после исключения из нее служебных слов. Столбцы в таблицах распределяются следующим образом:

- 1) нумерация ЛЕ языка L_i ;
- 2) ЛЕ языка L_i с их грамматическими пометами;
- 3) частоты ЛЕ языка L_i ;
- 4) нумерация ЛЕ языка L_j ;
- 5) ЛЕ языка L_j с их грамматическими пометами;
- 6) частоты ЛЕ языка L_j ;
- 7) разность между номером ЛЕ языка L_i и номером соответствующей ей ЛЕ языка L_j на данной строке;
- 8) квадрат разности по каждой строке; внизу – сумма всех квадратов разностей в столбце.

По полученным результатам построена сводная таблица, в которой к каждой рассматриваемой паре языков приводятся коэффициенты их корреляции.

Из нечеткого множества степеней сходства при помощи операции сгущения мы получили три:

- 1) полное сходство; сюда входят случаи полного соответствия одновременно по форме и по основным значениям, причем допускается формальное расхождение по фонетическим и/или морфологическим за-

конам (так, в паре исп. *estar* – ит. *stare* фонетически закономерно соответствие *es – s* в начале слова перед согласным и морфологически закономерно соответствие *-r – -re* в инфинитиве), а также отнесение одной и той же лексической единицы в частотных словарях разных языков к разным частям речи без изменения ее значения (так, в ЧС испанского языка слово *uno* отнесено к именам числительным, а в ЧС румынского языка соответствующее ему *unu(l)* помечено как местоимение);

2) частичное сходство; сюда входят случаи расхождения, лишь отчасти обусловленного фонетическими и/или морфологическими законами (ср.: исп. *cada* – фр. *chaque*), префиксального и постфиксального расхождения (ср.: исп. *entonces* – рум. *atunci*; исп. *aqun* – ит. *qui*; исп. *cómo* – фр. *comment*), а также вхождения слова в одном языке в состав эквивалентного ему сложного слова в другом (ср.: исп. *hoy* – фр. *aujourd'hui*);

3) несходство; сюда входят случаи полного расхождения по форме при полном или частичном совпадении основных значений, а также совпадения аффиксов при полном несходстве корневых морфем (ср.: исп. *encontrar* – рум. *întâlni (întâlni)*).

В случаях супплетивизма работа велась следующим образом. В алфавитных списках ЧС приводятся частоты каждой формы того или иного слова. Это позволяет разбить супплетивные формы таким образом, чтобы сопоставить их сходные в языке L_i и языке L_j пары по отдельности. Возьмем, например, испанский глагол *ser* и эквивалентный ему французский глагол *être*. Для начала следует выделить ряды их форм по общим основам. В списке форм исп. *ser* выделяется четыре таких ряда: *er-*, *es* (одна форма), *fu-*, *s-*; сходным образом – у фр. *être*: *es + est*, *ét- + et-*, *fu-*, *s-*. При выделении каждого такого ряда производится сложение частот тех форм, которые к нему относятся. Выделив эти ряды, мы получаем возможность сопоставить их по отдельности. Так, в полные соответствия включаются: исп. *es*

и фр. *es + est*; исп. *fu-* и фр. *fu-*; исп. *s-* и фр. *s-*. Что касается фр. *ét- + et-*, то этот ряд форм по фонетическим ключам входит в полное соответствие с исп. *estar*. При всех этих операциях происходит соответствующее перераспределение номеров в частотном списке как языка L_i , так и языка L_j , но лишь относительно каждой данной пары языков.

Определенные проблемы в приведенном примере вызывает исп. *er-*, не входящее в сопоставление ни с одним рядом французских форм. Ввиду этого исп. *er-* исключается из рассмотрения. Однако в том случае, если такие «лишние» основы выявляются и в языке L_i , и в языке L_j , они приводятся в таблице полных несоответствий. При этом, чтобы избежать излишней разбивки ЛЕ, основы, относящиеся к одной и той же ЛЕ, даются в таком случае на одной строке вне зависимости от того, однородны они или разнородны. Например, у испанского глагола *ir* выделяются следующие группы форм: *fu-*, *i- + yendo*, *v-*. Последняя входит в полное соответствие с основой *v-* эквивалентного итальянского глагола *andare*. В то же время две другие группы форм исп. *ir*, так же как основа *and-* ит. *andare*, остаются «незадействованными». Поэтому они заносятся в одну строку таблицы с полными несоответствиями, что выглядит следующим образом:

55.	ir v. (fu- + i- + yendo)	815	49.	andare v. (and-)	1115
-----	--------------------------	-----	-----	------------------	------

Второстепенные значения в данной работе не учитываются. Ведущим критерием выделения основных и второстепенных значений является частотность: если какой-либо ЛЕ языка L_i в пределах области полных и частичных соответствий эквивалентны две (и более) ЛЕ языка L_j , одна из которых по своей частотности меньше данной ЛЕ языка L_i в десять раз и более, то такая ЛЕ языка L_j исключается из рассмотрения (операция сгущения). Ср.: исп. *verdad* n. 326 – фр. *vérité* n. 97 + *vrai* n. 27. В данном случае исключается фр. *vrai*, частотность которого более чем в десять раз меньше частот-

ности исп. *verdad*. В то же время если соответствие с подобным соотношением частот является единственным, то оно учитывается (ср.: исп. *cuando* av. 803 – фр. *quand* av. 19).

В тех случаях, когда ЛЕ языка L_i имеет в языке L_j полные и/или неполные соответствия, а также соответствия с полным расхождением по форме, последние исключаются из рассмотрения. Например, исп. *palabra* имеет французские эквиваленты *parole* и *mot*. С первым из них оно обнаруживает полное сходство, со вторым – полное несходство. Соответствие исп. *palabra* – фр. *mot* не рассматривается. Вместе с тем действует правило основного значения (см. выше). Ср.: исп. *ciudad* п. 318 – фр. *ville* п. 184 + *citii* п. 24. Со словом *ville* испанское *ciudad* обнаруживает полное несходство, со словом *citii* – полное сходство. Однако *citii* имеет недостаточно высокую частотность по сравнению с *ciudad*, в отличие от *ville*. Поэтому соответствие исп. *ciudad* – фр. *ville* учитывается, а соответствие исп. *ciudad* – фр. *citii* исключается из рассмотрения.

Сходства, восходящие не к латинско-романскому, а к индоевропейскому периоду, в данной работе считаются несходствами. Так, например, обстоит дело с парой исп. *ciento* – рум. *sută*. Если испанское слово в этой паре является исконно романским, то румынское слово имеет славянское (по другой версии – иранское²) происхождение. В связи с этим данная пара относится к несходствам, несмотря на то что оба слова в ней восходят к одной и той же индоевропейской основе. Другой пример: исп. *luz* – фр. *lumière*. Их латинскими этимонами являются, соответственно, *lux* (основа *luc-*) и *luminaria* (< *lumen*). Слово *lux*, в свою очередь, происходит от и.-е. **louk-s*, а *lumen* – от родственного ему и.-е. **leuqs-men* (или **louqs-men*)³. В латинский период эти два слова не воспринимались как однокоренные, так как имели уже разные основы. Поэтому указанная испанско-французская лексическая пара заносится в таблицу несходств.

Обработка результатов, представленных в таблицах, была ориентирована на коэффициент корреляции по Спирмэну:

$$\rho = 1 - \frac{6 \sum_{i=1}^{i=n} d^2}{(n^3 - n)} = 1 - \frac{6 \sum_{i=1}^{i=n} d^2}{n(n^2 - 1)},$$

где d – разность между номером ЛЕ языка L_i и номером соответствующей ЛЕ языка L_j , n – количество строк с лексическими сопоставлениями в данной таблице, i – номер в ЧС.

С учетом трех перечисленных выше степеней сходства приходилось усложнять математическую обработку, используя формулу Спирмэна в измененном виде:

$$\rho = \left(1 - \frac{6 \left(\sum_{i=1}^{i=n} d_1^2 + 0,5 \sum_{i=1}^{i=n} d_2^2 \right)}{(n_1 + n_2)((n_1 + n_2)^2 - 1)} \right) \cdot \frac{n_1 + n_2}{n_1 + n_2 + n_3},$$

где d_1 – разность между номером ЛЕ языка L_i и номером соответствующей ЛЕ языка L_j в таблице с полными сходствами, d_2 – разность между номером ЛЕ языка L_i и номером соответствующей ЛЕ языка L_j в таблице с частичными сходствами, n_1 – количество строк с лексическими сопоставлениями в таблице с полными сходствами, n_2 – количество строк с лексическими сопоставлениями в таблице с частичными сходствами, n_3 – количество строк с лексическими сопоставлениями в таблице с несходствами.

Приведем пример вычисления по результатам таблиц при помощи формулы Спирмэна:

Исп.-фр. (по номерам ЧС знаменательных ЛЕ):

$$c = (1 - 6(147287 + 0,5 \cdot 31115) : (103 + 22) \cdot ((103 + 22)^2 - 1)) : ((103 + 22) : (103 + 22 + 55)) \approx (1 - 6 \cdot 162844,5 : 1953000) \cdot 0,7 \approx 0,5 \cdot 0,7 = 0,35$$

Таблица 1

Результаты

Пара	Коэффициент корреляции	
	По номерам общего ЧС	По номерам ЧС знаменательных ЛЕ
Испанский–французский	0,15	0,35
Испанский–итальянский	0,21	0,4
Испанский–румынский	-0,03	0,18
Французский–итальянский	0,3	0,56
Французский–румынский	-0,69	-0,12
Итальянский–румынский	-0,4	-0,08

Исходя из полученных коэффициентов корреляции, можно выстроить следующую иерархию исследуемых пар языков по убывающей степени близости: французский–итальянский, испанский–итальянский, испанский–французский, испанский–румынский, итальянский–румынский, французский–румынский. В таком случае относительно западнороманских (испанского и французского) языков итальянский язык представляется центральным (каковым он, собственно говоря, и является – по крайней мере, с историко-географической точки зрения) с «иррадиацией» в ареалы других романских языков. Эти ареалы при этом более удалены друг от друга, чем относительно центра, с которым они в языковом плане обнаруживают различную степень схождения. Такая схема соотносится с результатами исследования Дж.Э. Ри⁴.

Следует в то же время обратить внимание на отрицательные коэффициенты

корреляции в парах с румынским языком, свидетельствующие о значительной структурно-синергетической отдаленности румынского от других романских языков и опровергающие выдвигаемую нами ранее гипотезу, согласно которой между родственными языками возможен показатель от 0 до +1. Эти данные можно интерпретировать в пользу концепции Западной и Восточной Романии, однако итальянский язык тогда будет примыкать к Западной Романии, а не к Восточной, вопреки классической схеме⁵. При этом наибольшая (хотя также достаточно низкая) степень структурно-синергетической близости к нему обнаруживается у испанского языка. Этот факт отчасти оправдывает деление Романии на центральную и периферийную⁶. В целом же получается, что наиболее типичным романским языком является итальянский, наименее типичным – румынский⁷.

ПРИМЕЧАНИЯ

¹ Бабурин Л. К. Степень близости испанского языка с французским и с итальянским // Известия Российского государственного педагогического университета им. А. И. Герцена: Аспирантские тетради, 2007. № 15 (39). С. 42–50; *Он же*. Степень взаимной системно-синергетической близости испанского, французского и итальянского языков // Известия Российского государственного педагогического университета им. А. И. Герцена: Аспирантские тетради. 2008. № 27 (61). С. 31–34.

² Ср.: *Cioranescu A.* Diccionario etimológico Rumano: en 7 fasc. Madrid: Biblioteca Filológica, Universidad de La Laguna, 1958–1966. P. 814.

³ *Walde A.* Lateinisches etymologisches Wörterbuch. 3., neubearbeitete Aufl. von J.B. Hofmann. Heidelberg: Carl Winter's Universitätsbuchhandlung. 1. Bd. (A-L), 1938. S. 824, 832.

⁴ Ср.: *Kroeber A.L.* Three Quantitative Classifications of Romance // Romance Philology. Berkeley, Los Angeles: University of California Press, Vol. XIV. N 3, 1961. P. 193.

⁵ Ср.: *Diez F.* Grammatik der romanischen Sprachen. 1. Teil. 5. Aufl. Bonn: Eduard Weber's Verlag, 1882. S. 1; *Fuchs A.* Die romanischen Sprachen in ihrem Verhältnisse zum Lateinischen. Halle: H.W. Schmidt, 1849; *Wartburg W. von.* Die Entstehung der romanischen Völker. Halle, Saale: Max Niemeyer Verlag, 1939.

Ср. сходные выводы в: *Hall Jr. R.A.* The Reconstruction of Proto-Romance // Language. Journal of the Linguistic Society of America. Baltimore: Waverly Press, Inc., Vol. 26. N 1, 1950. P. 6–27; *Mańczak W.* Problemy językoznawstwa ogólnego. Wrocław, Warszawa, Kraków: Ossolineum, 1996. S. 177.

⁶ Ср.: *Klein F.-J.* Bedeutungswandel und Sprachendifferenzierung: Die Entstehung der romanischen Sprachen aus wortsemantischer Sicht. Tübingen: Max Niemeyer Verlag, 1997.

⁷ Ср.: *Mańczak W.* Op. cit. S. 177.