Yuri Tambovtsev
Novosibirsk Pedagogical University, Russia

Consonant patterns in Spanish and other Romance languages: A study in typological closeness

Spanish is said to belong to the group of Romance languages in the Indo-European family. Besides Spanish, the Romance group includes Portuguese, French, Italian, Rumanian, Moldavian and some other minor Romance languages. Many linguists believe that the substratum language for Spanish was the language of the Iberians who lived in their native territory before they were conquered by the Romans. After that, however, they were conquered by German tribes in the 5th century and the Arabs in the 16th century. One can find lexical borrowing both from the ancient German and Arab. However, in our view these invasions did not change the articulation base of the ancient Spanish language. The Castilian dialect, which was the basis for the literary Spanish language, is said to be quite different from Latin (Tambovtsev, 2001-a; 2001-b).

The typological closeness of languages may be measured by a number of features. By the closeness we mean the distances between languages. The goal of this article is to measure the phono-typological distances between Spanish and the other five major Romance languages in question, i.e. Portuguese, French, Italian, Rumanian and Moldavian by the total of the distribution of the sounds in the language speech chain.

Likewise, since Latin is the parent language of the Romance languages, it is advisable to compare its speech sound pattern to Spanish. Esperanto should also be compared to Spanish since it is built mainly on the linguistic material of the Romance languages, though Esperanto is an artificial language. Some linguists may say it is not possible to compare an artificial language with a natural language. Let us consider if Esperanto is really an artificial language on the phonetic level. It is quite artificial on the lexical and syntax level, as different word

stocks are mixed, different affixes are taken from different languages and glued together. However, on the phonetic level Esperanto is in no way different from any natural language. If on the lexical level it has hybrid words, on the phonetic level there are no hybrids. Esperanto uses natural speech sounds, i.e. phonemes. Thus, in this respect, it is quite a natural language. Really, every linguist is under a false impression that Esperanto is an artificial language on the phonetic level. If a linguist analyses Esperanto phonemes, he comes to the conclusion that the constructions of its phonemes are not hybrids. Its speech sounds are the same as the speech sounds in the natural languages on whose basis Esperanto has been built.

Thus, from the point of view of phonetics Esperanto is the same sort of object as any other natural language. It has the same phonetic (articulatory) features as a natural language. Esperanto's speech sounds may be classified in the same way as the speech sounds of any natural language according to the work of the organs of speech, the manner of articulations, the work of the vocal cords, etc. (Tambovtsev, 2001-a: 98 - 102). This principle also concerns the other artificial languages. They, too, are built with the help of natural human speech sounds. Nevertheless, Esperanto, Glossa, Ido, Interlingua, Novial, and Volapuek are called artificial languages (Crystal, 1992: 28-29) because they are constructed as hybrids on the lexical level. Thus, the word structure plays the predominant role in their classification. However, from the point of view of phonetics, all these artificial languages may be compared to natural languages since they can be pronounced. If they can be pronounced, they have the same sets of phonetic (articulatory) features. (Tambovtsev, 2003).

Every language has a unique structure of distributions of speech sounds in its phonemic chain (Tambovtsev, 2002-a; 2002-b), and Spanish, like any other human language, has a specific structure of its speech sound chain. It can be distinguished by its structure from any other language. World languages may be classified in language groups by different means. We applied the phono-typological distances as the means for their classification. However, it

would be interesting to compare the results of our classification to those which are generally accepted. In fact, the most common classification of languages is the genetic classification.

The genetic classification of world languages by the classical comparative method is widely accepted. Nevertheless, April and Robert McMahon are quite correct in saying that there are various methods for classifying languages, which yield different results, i.e., different classifications. All of them have some drawbacks and some are extremely controversial. They give the Greenberg method of mass comparison as an example of the most controversial method, i.e., which gives the most controversial result of classification (McMahon et al., 2005: 5). At the same time, they point to the fact that the well-known comparative method fails to reconstruct Latin on the basis of the Romance languages, which are the daughter languages of Latin (McMahon et al., 2005:14-19). One should agree with them that it is high time to improve the comparative method or to invent a new, more reliable method.

Our method may be another good method which can help to classify languages into language taxa. It does so according to their sound similarity. It gives formal and repeatable results, which the comparative method cannot give. Stating that the comparative method is rather subjective, the McMahons believe that this brings the comparative method uncomfortably close to Greenberg's method of mass comparison, where accepting results means believing implicitly in the linguistic intuitions of the method inventor. Both methods rely on an individual linguist's knowledge of a particular language group, but this makes both inevitably subject to inference from the individual linguist's opinion. One cannot help agreeing with them that if we cannot guarantee getting the same results from the same data considered by different linguists, we jeopardize the essential scientific criterion of repeatability (McMahon, et al., 2005: 27).

We must emphasize again that McMahon, et al., are quite correct to point to the fact that

the well-known comparative method fails to reconstruct Latin on the basis of the Romance languages, though we know that they are the daughter languages of Latin (McMahon, et al., 2005:14-19). Still there is no answer to this question: should all the Indo-European languages be considered one language family or should it be divided into several language families? So, in the other words, should it be considered a language family or a language union? The languages which enter this language taxon are really very different (Tambovtsev, 1986; 1989).

One should agree with McMahon, et al., that it is high time to improve the comparative method or to invent a new, more reliable method. Our method may be another good method which can help to classify languages into language taxa. It does so according to their sound similarity. It gives formal and repeatable results, which the comparative method cannot give. Stating that the comparative method is rather subjective, McMahon, et al., believe that this brings the comparative method uncomfortably close to Greenberg's method of mass comparison, where accepting results means believing implicitly in the linguistic intuitions of the method inventor. Both methods rely on an individual linguist's knowledge of a particular language group, but this makes both inevitably subject to inference from the individual linguist's opinion. One cannot help agreeing to them that if we cannot guarantee getting the same results from the same data considered by different linguists, we jeopardize the essential scientific criterion of repeatability (McMahon et al., 2005: 27).

The criterion of repeatability is quite important in the Natural Sciences. It allows observers to be exactly objective. If we want to obtain reliable results we must understand that our results may be repeated by other linguists. The history of the Natural Sciences is a good example of proving that only the objective results can be fundamental because they can be repeated by any other scholar (Kuhn, 1977:22). Unfortunately, T.S. Kuhn is correct to remark that it is hard for new methods to break through. It is especially true for the

Humanities, in which scientific paradigms are not verified for centuries. Our method of establishing distances may verify some of the fundamental theories in linguistics but it can also show that some of the methods are too subjective and some theories do not hold. It is high time to introduce the data on language distances into linguistics which may be a turning point to reconsider some of its classification which are more geographical than linguistic.

It is especially interesting to try our method on some Romance language since we have a strong evidence of the parent language – Latin. We have chosen Spanish as a reference point for the simple reason that we know that it acquired its stock of words from Latin. Therefore, the sound picture of Latin should be considered by all means. So, here, Latin is a reference point.

An artificial language, Esperanto, has been taken as another reference point, the reason for which is that we know the source from which it obtained its basic stock of words. Thus, this artificial language may help us to determine how well our method works. L. L. Zamenhof, a doctor from Poland, created Esperanto in 1887. He had a nickname *Esperanto* which means "Hoping" in Latin. His nickname became the official name of this artificial language. The language of Esperanto comprises the roots of some European languages: Latin, Old Greek, Greek, Italian, Spanish, French, English, Russian, and Polish. These roots serve as the base to create comprehensive speech with the help of some affixes, which transfer grammatical meanings (Tambovtsev, 2001-a: 98).

It seemed to us really interesting to see to which Romance language(s) Esperanto would be close. One should also take into account that the sound chain of Esperanto may be unique because it uses only several fixed affixes. Therefore, its sound chain in speech may be quite different from any other natural language. As we can see from the phono-typological distances provided below (p. 7), Spanish is the closest to Esperanto, but not Latin, from which Spanish originated.

The new method is based on both phonostatistics and metric analysis in order to put it in this or that language taxon. The new linguistic method of metric measurements is used here. It was first proposed some time ago (Tambovtsev, 1977), but it was never used by other linguists. It was developed further by the author later (Tambovtsev, 1994-a; 1994-b).

The distribution of Spanish vowels will not be considered till the second stage of the investigation. Let's point out that consonants bear the semantic load in the word, not vowels. Therefore, it is more possible to understand the meaning of the message by consonants than by vowels. However, if we fail to recognise and distinguish two languages, then we resort to the structure of occurrence of vowels in the speech sound chain. While comparing languages, it is necessary to keep to the principle of commensurability. Bearing that in mind, it is not possible to compare languages on the basis of the frequency of occurrence of separate phonemes, because the sets of phonemes in languages are usually different.

The articulatory features may serve as the basic features in phono-typological reasoning. First of all, it is the classification of consonants according to the work of the active organ of speech or place of articulation (4 features: labial, front, palatal, and velar). Secondly, it is the classification from the point of view of the manner of articulation or the type of the obstruction (3 features: sonorant, occlusive and fricative). Thirdly, it is the classification according to the work of the vocal cords (1 feature: voiced). In this way, 8 basic features are obtained: 1) labial; 2) front; 3) mediolingual or palatal; 4) back or velar; 5) sonorant; 6) occlusive; 7) fricative; and 8) voiced consonants. One should take the values of the frequency of occurrence of these 8 features in the speech chain of Spanish and compare them to those of the other Romance languages. On the basis of the "chi-square" test and Euclidean distance, we have developed our own method of measuring the phono-typological distances between languages. It is described in detail elsewhere (Tambovtsev, 2003). It takes into account the frequency of occurrence of the 8 consonantal groups mentioned above to measure the

overwhelming mosaic of the language sound picture.

Having compared Spanish to some Romance languages (c.f. Table1), we received the following phono-typological distances:

Spanish – Moldavian 6.22
Spanish – Rumanian 7.75
Spanish – Portuguese 7.91
Spanish – Latin 7.98
Spanish – Esperanto 8.19
Spanish – Italian 8.35
Spanish – French 13.39

The distances from Spanish to other varieties in the Romance taxon, excluding French, cluster closely around the mean of the calculated distances (8.79). French falls distant from the mean by more than 4.5 units; this is even more striking if one excludes the distance calculated for Spanish – French, which gives a mean between Spanish and the other varieties of (7.87). We notice how closely the distance from Spanish to Rumanian and Spanish to Portuguese lie from the mean that excludes French. Even then, Spanish – Moldavian is least distant, more than 1 unit from the mean.

As a conclusion, we can state that the speech sound picture of Spanish is most similar to Moldavian. One can account for this by the fact that both Moldavian and Spanish use many Latin roots. It is known that both Spanish and Moldavian were influenced by the same sort of substratum. We can suppose that the Spanish substratum changed the Latin speech sound patterns in the same way as the Moldavian substratum did.

It is not a surprise that Spanish is close to the sound pattern of Portuguese or Latin. In fact, the distance between Spanish and Portuguese and between Spanish and Latin are within 0.1 of the mean calculated (excluding French). Actually, Latin is the parent language of all Romance languages, thus Spanish uses many of its roots. Spanish is rather far away from the language which is its geographically close neighbor, i.e. French, which has undergone many

phonetic changes because of its substratum. The articulation basis of the French substratum language was quite different from that of the Spanish substratum language. Therefore, it is logical that Spanish and French are distant from each other by the phonostatistical and typological points of view. The sound pattern of French, as we have demonstrated, shows a greater distance from Spanish than the rest of the Romance languages from Spanish.

Our data state that the speech sound pattern of Spanish more or less resembles that of the other Romance languages. Latin, the parent language for all the Romance languages, gave all Romance languages but French its basic sound system. We must point out, therefore, that it is not a coincidence that Portuguese, Italian and Romanian are more similar to Spanish than French. It was no surprise for us that Spanish consonants are distributed more similar to Moldavian, Portuguese, and Italian, because of the similar articulation bases of their substratum languages.

Tab.1 The Frequency of occurrence of phonemes in Spanish and some other languages

The samples in phonemes: Spanish (SPN) – 500 000 (Lloyd et. al., 1967); Portuguese (PRTG) – 5000 (Zipf et al., 1939); Italian (ITLN) – 103422 (Boldrini, 1948); Rumainian (RMN) – 38038 (Tambovtsev, 2001); Moldavian (MLDV) – 40021 (Tambovtsev, 2001); French – 2122962 (Chavasse, 1948; De Kock, 1982; Haton et. al., 1971; Hug, 1979; Szklarczyk, 1961; Veronis, 1983; Wioland, 1974); Latin (LATN) – 391121 (Tambovtsev, 2001); Esperanto (ESPR) – 3580031 (Tambovtsev, 2001); Greek (GREK) – 17690 (Mirabel, 1959); Armenian – 41987 (Melkumian, 1967).

The frequency of occurrence in per cent (%) to all the sounds in the speech sound chain:

№	Phoneme Class	SPNS	PRTG	ITLN	RMN	MLDV	FREN	LATN	ESPR	GREK	ARMN
1.	Labial	9.79	11.10	10.38	9.20	11.06	13.96	12.29	9.03	10.81	10.32
2.	Front	33.09	35.44	38.46	36.70	35.77	25.43	37.88	36.85	33.39	36.24
3.	Palatal	1.09	4.09	1.77	2.29	2.68	2.20	1.70	2.92	2.55	2.87
4.	Guttural	6.63	4.63	4.86	6.61	5.84	15.10	6.74	5.32	7.02	9.98
5.	Sonorant	19.35	22.67	24.99	24.71	22.89	24.06	23.77	25.47	20.71	27.75
6.	Occlusive	19.44	17.99	20.60	21.54	22.64	19.14	22.92	18.28	17.09	24.66
7.	Fricative	11.81	14.60	9.88	8.55	9.82	13.49	11.92	10.37	15.96	7.00
8.	Voiced	8.70	13.32	8.67	7.48	8.46	10.95	7.14	6.16	5.86	6.77
9.	Vowel	49.40	44.74	44.53	45.20	44.65	43.31	41.39	4588	46.23	40.59

References

Crystal, 1992 - Crystal, David. *An Ecyclopedic Dictionary of Language and Languages*. - Oxford: Blackwell, 1992. - 428 p.

Baudot, 1968 – Baudot J.A. *Information, redodance et repartition des letters et des phonemes en francais.* - Montreal: Universite de Montreal, 1968. - p.74.

Boldrini, 1948 - Boldrini M. *Le statistiche letterarie e i fonemi elementari nella poesia*. - Milano, 1948.

Chavasse, 1948 – Chavasse P. *Essai sur la phonetique statistique de la langue francaise et son application a l'etude de l'intelligibilite d'une conversation*. – Annales des Telecommunications. Cahiers d'acoustique, 1948, t.3, N1, p.5-23.

Haton et al., 1971 – Haton J.-P., Lamotte M. *Etude statistique des phonemes et diphonemes dans le francais parle.* – Revue d'Acoustique, N 16, 1971, p.258-262.

Hug, 1979 - Hug M. *La distribution des phonemes en Francais. Die Phonemverteilung im Deutschen.* – Geneve: Slatkine, 1979. – 192.

Kuhn, 1977 – Kuhn T.S. *The Structure of Scientific Revolutions*. - Moscow: Progress Publishing House, 1977. - 300 pages [in Russian].

Lafon, 1961 – Lafon J. Message et Phonetigue. – Paris: P.U.F., 1961. – p.142-146.

Lloyd et al., 1967 – Lloyd, Paul M. and Schnitzer, Ronald D. *A statistical study of the structure of the Spanish syllable.* - In: *Linguistics, #37*, 1967, p.58 – 72.

McMahon et al., 2005 - McMahon, April and McMahon, Robert. *Language Classification by Numbers*. - Oxford: Oxford University Press, 2005. - 263 pages.

Melkumian, 1965 – Melkumian M.R. Iskhodnye dannye i statistiko-kombinatornoe vydelenie paradigmy pervogo morfologicheskogo tipa v armianskom jazyke. [Raw data and statistic-combinational definition of the paradigm of the first morphological type in the Armenian language – Statistiko-Kombinatornye Metody v Jazykovedenii [Statistic-Combinatoric Methods in Linguistics], 1965, p.123 - 136.

Mirambel, 1959 - Mirambel A. *La langue grecque moderne: description et analyse.* - Paris: Klincksiek, 1959. – 157.

Szklarczyk, 1961 – Szklarczyk L. Essai sur la structure phonologique du francais. Ph.D. University of Pennsylvania. Philadelphia: University of Pensylvania, 1961.

Tambovtsev, 1977 - Tambovtsev, Yuri. Nekotorye harakteristiki raspredelenija fonem mansijskogo jazyka [Some Characteristics of the Distribution of the Mansi Language]. - In: Soviet Finno-Ugric Studies, Volume XIII, # 3, 1977 (Tallin), p. 195 - 198.

Tambovtsev, 1986 - Tambovtsev, Yuri. Konsonantnyj koeffitsient v jazykah raznyh semej [Consonant Coefficient in the Languages of Different Families]. - Novosibirsk: Novosibirsk University, 1986. - 17 pages.

Tambovtsev, 1989 - Tambovtsev, Yuri. Review on the book by Tomas V. Gamkrelidze and Vjach. Vs. Ivanov. Indo-European and the Indo-Europeans. Tbilisi: The Tbilisi University Press, 1984. Volume 1 and 2. - In: Gengo Kenkyu, # 96, 1989 (Tokyo), p. 119-143.

Tambovtsev, 1994-a - Tambovtsev, Yuri. Dinamika funktsionorovanija fonem v zvukovyh tsepochkah jazykov razlichnogo stroja. [Dynamics of functioning of phonemes in sound chains of the languages of different structure]. - Novosibirsk : Novosibirsk University.

Tambovtsev, 1994-b - Tambovtsev, Yuri. Tipologija uporjadochennosti zvukovyh tsepej v jazyke.[Typology of orderliness of the sound chains in language]. - Novosibirsk: Novosibirsk University.

Tambovtsev, 2001-a - Tambovtsev, Yuri. Kompendium osnovnyh statisticheskih harakteristik funktsionirovanija soglasnyh fonem v zvukovoj tsepochke anglijskogo, nemetskogo, frantsuskogo i drugih indoevropejskih jazykov. [Compendium of the basic statistical characteristics of functioning of consonants in the speech chain of English, German, French and other Indo-European languages]. - Novosibirsk: Novosibirskij klassicheskij institut, 2001. - 129 pages.

Tambovtsev, 2001-b - Tambovtsev, Yuri. Perednejazychnye soglasnye kak pokazatel' odnoj iz lingvisticheskih universalij jazykov mira [Front consonants as an indicator for one of the universals of world languages]. - In: Sibirskij Lingvisticheskij Seminar, # 2 (2), Novosibirsk, 2001, p.18 - 27.

Tambovtsev, 2002-a - Tambovtsev, Yuri. Comparative typological study of language distances based on the consonants in sound chains of various languages. - In: *The 5th National Colloquium for Computational Linguistics in the UK. Proceedings of the Conference*. (Edited by John Elliot). 8-9 January, 2002. University of Leeds, UK. - Leeds: University of Leeds, 2002-a, p.77-80.

Tambovtsev, 2002-b - Tambovtsev, Yuri. Is Kumandin a Turkic language? - In: Dilbilim Arashtirmalari (Istanbul), 2002-b, p. 63 - 104.

Tambovtsev, 2003 - Tambovtsev, Yuri. Tipologija Funktsionorovanija fonem v zvukovoj tsepochke indoevropejskih, paleoaziatskih, uralo-altajskih i drugih jazykov mira: kompaktnost' podgrupp, grupp, semej i drugih jazykovyh taksonov. [Typology of functioning of phonemes in sound chain of Indo-European, Paleo-Asiatic, Ural-Altaic and other world languages: compactness of subgroups, groups, families and other language taxins]. - Novosibirsk: Sibirskij nezavisimyj institut, 2003-a. - 143 pages. [in Russian].

Tambovtsev, 2008 - Tambovtsev, Yuri. Typological Proximity of Esperanto to the Classical language Taxa of the Indo-European Family. - In: California Linguistic Notes, Vol. 33, #1, Winter, 2008, p.1-51.

Veronis, 1983 – Veronis, Jan. La difficulte dans les relations phonie-graphie: une etude quantitative. - In:Travaux de l'institut de phonetique d'Aix, Vol.9, 1983-1984, p.241-263.

Wioland, 1974 – Wioland F. Contribution a l'etablissement de constantes en relation avec la frequence des phonemes en francais parle. - Travaux de l'institut de phonetique de Strasbourg, N 6, 1974, p.141-164.

Zipf et al., 1939 – Zipf G.K. and Rogers F.M. Phonemes and Variphones in Four Present-Day Romance Languages and Classical Latin from the Viewpoint of Dynamic Philology. – In: Archives Neerlandaises de Phonetique Experimentale, 15, 1939, p.112-147.