THE LINK BETWEEN POLYGLOTTERY AND PERCEPTUAL DIALECTOLOGY: A CASE STUDY OF RUSSIAN DIALECTS

Yan Aleshkevich-Suslov

Moscow State University
1 bld. 13, Leninskie Gory, Moscow, 119991, Russia

Abstract. The status of language varieties as dialects or distinct languages has long been a controversial topic, as the distinction is often coloured not only by objectively measurable linguistic data, but also by history, speakers' attitudes, prejudices, metalinguistic awareness and general education. It is the author's belief that one could establish criteria for differentiating a language from a dialect by asking the speakers of a majority language directly in a survey-based blind test, thus accounting for the complex interaction of factors that affect language perception, but mitigating the biases of socio-cultural influences. This study proposes a simple method for checking one-way intelligibility in lieu of a mutual intelligibility test. Another issue that this article is concerned with is polyglots. Polyglots are known for being able to study languages efficiently, presumably, due to their increased language aptitude and awareness. Can this awareness have an effect on a polyglot's propensity for perceiving dialects as languages? Answering this question is the second task of this paper. The results of the study show a rather weak positive correlation between the number of languages that a person knows and his tendency to identify an unknown speech sample as a language rather than a dialect; however, they do not refute this idea outright. Additionally, the author found setting a criterion for differentiating a language from a dialect difficult due to an unexpectedly high intelligibility rate of a lect that was known to be a distinct language. This implies that further testing of this sort needs to be done. It was, however, established that speakers of Russian tend to see a dialect as a lect that only differs in phonetics, while a language, in their perception, is a lect that differs in phonetics as well as vocabulary.

Keywords: polyglottery, perceptual dialectology, languages and dialects, Russian dialects

СВЯЗЬ ПОЛИГЛОТИИ И ПЕРЦЕПТИВНОЙ ДИАЛЕКТОЛОГИИ НА МАТЕРИАЛЕ ДИАЛЕКТОВ РУССКОГО ЯЗЫКА

Я. В. Алешкевич-Суслов

Московский государственный университет им. М. В. Ломоносова
119991, Россия, Москва, Ленинские горы, 1, стр.13

Аннотация. Статус языковых разновидностей как диалектов или отдельных языков с давних пор является неоднозначным вопросом, так как диалекты разграничиваются на основании не только объективных лингвистических данных, но и на основании истории, отношения носителя к языковой разновидности, его предубеждений, металингвистических сведений и общего уровня образования. Автор полагает, что можно установить критерии разграничения языка и диалекта путём прямого опроса носителей доминирующего языка, в котором респонденты не будут знать, с какими вариантами языков они имеют дело. Таким образом можно учесть сложную взаимосвязь факторов, влияющих на восприятие языков, и одновременно снизить роль социокультурных воздействий. В настоящем исследовании предлагается простой метод проверки односторонней понятности вместо теста на взаимную понятность. Другой вопрос, который исследуется в данной статье, связан с полиглотами: предполагается, что полиглоты способны изучать языки эффективно благодаря повышенному чувству языка. Может ли это чувство влиять на их склонность воспринимать диалекты как языки? Ответ на этот вопрос является второй целью настоящей работы. Результаты исследования показывают лишь довольно слабую положительную корреляцию между числом языков, которые знает человек, и вероятностью того, что он назовёт языковой вариант языком, а не диалектом, однако не отвергают полностью выдвинутую гипотезу. Также автор пришёл к выводу, что критерий для разграничения языков и диалектов установить трудно ввиду неожиданно высокой понятности языковой разновидности, о которой известно, что она является отдельным языком. Это означает, что необходимо продолжение аналогичного опроса в более широком масштабе. Однако было установлено, что носители русского языка склонны воспринимать диалект как ту разновидность языка, которая отличается от родного им языка только своей фонологией, а язык – как разновидность, которая отличается и фонологией, и словарём.

Ключевые слова: полиглотия, перцептивная диалектология, языки и диалекты, диалекты русского языка

Introduction

Polyglottery is the phenomenon of an adult making a conscious effort to acquire multiple languages [6]. Polyglottery is believed to be reinforced by an enhanced language awareness in the individuals known as polyglots [5], which implies that polyglots may be more sensitive to the minutia of a given language, and thus be able to more accurately discern between a language and a dialect.

Perceptual dialectology is a branch of folk linguistics that studies the distribution of linguistic features as perceived by speakers without a linguistic background. The aim of perceptual dialectology is to find out how social factors affect the perception of dialects, both in terms of the speakers' conscious opinions about languages and their unconscious perception of dialectal features [8].

The key idea in this study differs from that of mainstream perceptual dialectology in that the author is concerned not so much with the distribution of dialectal features, but rather with the status of a lect as a dialect or a language. The distinction between language and dialect is not due to any intrinsic linguistic features of a lect, but rather due to a complex interplay between sociological and sociolinguistic features, such as, but not limited to: positive or negative biases and attitudes towards the peoples of a given region where the lect in question is spoken; the speakers' interest in other cultures; previous experience studying other cultures; biases arising from political agendas inherent in the schooling programmes; metalinguistic awareness; experience with various registers in one's own language, etc. [4]. As such, a reasonable proposition would be to create a criterion for designating a lect as a dialect or a distinct language based on a majority vote in a blind test by speakers of closely related languages. In other words, when viewing separate languages as *Abstandsprachen* (a term denoting lects so disparate that they are to be considered separate languages) [ibid.], the author uses the speakers of a language as the measuring device to measure the distance by asking them directly. One could then evaluate whether the number of languages that a speaker knows correlates with how frequently he identifies a lect as a language or a dialect.

As such, the goals of this investigation are twofold:

1) to establish a criterion for distinguishing a dialect from a language based on native speakers' subjective opinion;

2) to examine the influence that knowing multiple foreign languages has on the perception of the distinction between a language and a dialect.

Common approaches to distinguishing dialects and languages

Creating a criterion for distinguishing dialects and languages has long been a goal of many researchers, and the most common approaches found in literature are the following [9]:

1) measuring mutual intelligibility;

2) measuring the differences between two lects computationally.

As pointed out by Wichmann [ibid.], mutual intelligibility is often asymmetrical (for instance, speakers of Danish understand spoken Swedish much better than vice versa) [4], and, additionally, mutual intelligibility is very difficult to measure. Thus, Wichmann [9] proposed to use an algorithm based on phonetic lexical difference using normalised Levenstein distance (the number of permutations needed to change the word from one lect into its semantic counterpart in another lect), or LDN for short. They came out with a cut-off LDN of 0.51 (which is a conservative estimate, other possible cutoffs are discussed in the paper): if the LDN between two lects is greater than that, they are considered separate languages. By such a classification, Indonesian and Malay, Bosnian and Croatian, and Hindi and Urdu constitute pairs of dialects of the same language, while Catalan and Spanish, Cairo Arabic and Moroccan Arabic, and Japanese and Miyako constitute pairs of closely related yet distinct languages. Such results are uncontroversial and are in line with both a public understanding and a typical linguistic classification of those lects, but some more unusual conclusions from this study are that Danish and Swedish, as well as Russian and Belarusian should be classified as pairs of dialects, albeit they are very close to the language threshold (LDN = 0.49 and 0.46, respectively).

While the findings of the paper mentioned above are promising and merit further investigation, the author believes that socio-cultural factors play a major role in differentiating languages, and an entirely computational method is incapable of accounting for that. Speakers' input must be considered, and while mutual intelligibility tests are indeed difficult to conduct, one-way intelligibility tests involving speakers of a majority language might provide a sufficient alternative, because when taking the democratic approach (everyone's opinions are treated equally), the influence of the speakers of the majority language would outweigh the speakers of the minority language. So, to a first approximation, gauging intelligibility from the viewpoint of the speakers of a majority language should suffice for a criterion to distinguish between a language and a dialect.

The first attempts to develop a methodology for testing dialect intelligibility were made by structuralists in the 1950s America, who focused on establishing linguistic borders in the Native American languages and devised the recorded text testing (RTT) to that end. This method has found use for literacy programmes to standardise orthographies, as well as language policies and sociolinguistic studies [4]. In this study, the author uses a somewhat modified approach similar to RTT.

The contribution of perceptual dialectology to the issue of dialect vs. language

Lee [7] proposes that even non-linguistic features, such as the amount of pauses and bad starts in a recorded speech sample, can influence a respondent's perception of a dialect, and points out that our current methods may be, as of yet, unrefined, and that the most reasonable attitude is that of healthy skepticism. Rather than attempt to further refine precise methods, the author, instead, turns to a simpler method of appealing to the masses to define a criterion for a language vs. a dialect.

According to Clopper [2], perceptual dialectology is concerned with issues such as how sociolinguistic representations are acquired throughout an individual's life span and how dialect variation and classification influence speech processing, and it aims to build a model of linguistic and social representations that influence dialect perception. One of the critical features of the author's proposed model for determining a speaker's aptitude for dialect perception is accounting for the speaker's experience, which includes both language acquisition in childhood and his continued learning throughout his life. The second critical design feature of the model is accounting for the effects of socio-cultural stereotypes, listener's expectations, and response biases.

In the same manner, the author wishes to find out whether experience with language of all sorts has any influence on dialect perception, that is, whether a polyglot's experience with a multitude of languages is correlated with an increase or a decrease in the number of languages that he identifies across a sample of closely related lects. The socio-cultural stereotypes are another important factor to consider, and one of the most obvious factors allowing speakers to identify a different language is its alphabet. Thus, in this study, which focused on written speech samples, the author sought to eliminate the differences in orthography.

Williams [10] stresses that the most important aspects for differentiating a language from a dialect are language attitudes, language use patterns, multilingualism, and the sociological aspects of a given region. To gauge the distance between the Cangin languages (a group of related languages in Africa), he used a combination of approaches: the phonostatistic method (comparing the number of differences in the phonetic features of an identical gloss, where a value higher than 100 represents a potential obstacle to comprehension) and a mutual intelligibility test. He found that the results of both the phonostatistics and mutual intelligibility tests displayed values similar to the ones accepted for considering two lects to be dialects of a single language. He thus concluded that the Cangin language group is composed of five distinct languages, contrary to what was previously believed. His paper demonstrates an example of a multi-faceted study that can be used for further similar studies in other linguistic areas where the linguistic status of some lects is highly disputed.

The questionnaire

The anonymous questionnaire on Google Forms[1] included a section where the respondents would provide their background data: their origin, education, an interest in linguistics, and, most importantly, the number of languages that they know. After that, the respondents were given 10 texts in sequence: each text was written in a normalized phonetic Cyrillic orthography (so as to eliminate potential differences in the respondents' perception of an unknown alphabet, as well as to facilitate vocal reproduction of the texts). The rules of the normalised orthography accompanied each text. Eight of the texts were recorded speech samples of the lects that are traditionally classified as dialects of Russian (texts were taken from [1], [3, p. 71–76], and online resources[2]), while two of the texts are excerpts from works written in literary East Slavic microlanguages, namely, Preshov Rusyn[3] and Podlachian[4]. The Smolensk lect was represented twice, owing to the lack of other suitable texts. The respondents were tasked with the following:

1. Reading the text out loud.
2. Rating the intelligibility of the text according to their subjective perception from one (lowest) to ten (highest).
3. Rating the comfort of conversation with a speaker of the lect presented in the text from one (lowest) to ten (highest).
4. Naming which aspects of the given lect (vocabulary, phonetics or grammar) differ the most from standard Russian.
5. Naming which aspects of the given lect (vocabulary, phonetics or grammar) are the closest to standard Russian.
6. Identifying the lect either as a dialect of Russian or as a distinct East Slavic language.

In the introduction to the questionnaire, the respondents were not told how many distinct East Slavic languages or how many dialects there were in the pool of texts. They were also not told which languages or dialects were featured in the questionnaire, so as to minimise the effects of psychological priming. Data manipulation was then carried out using the built-in features of Google Spreadsheets.

Participant profile

Responses were received from 153 participants. The target audience was teenagers and young adults. Over 80% of the respondents fell into the age group of 14 to 30 years, with an average of 24 years. Around 36% of the respondents work or study in the field of natural sciences, while around 24% work or study in the fields of literature and linguistics. Over half of the respondents were residents of Moscow and the greater Moscow area. Over half of the respondents claim to know one to three languages, while only about a fifth of the respondents can be considered polyglots knowing five or more languages (Fig. 1).

---

[1] Opros vo vospriiatiiu dialektov russkogo iazyka i vostochnoslavianskikh iazykov [Survey on the perception of Russian dialects and East Slavic languages], docs.google.com/forms/d/e/1FAIpQLSf3_tkS930ixKNZlYapKQ3tz7He53TbDvWQ3vN_jQxgV-_jqQ/viewform?usp=sf_link (accessed 10 August 2023).

[2] Galanin, A. Pomorskaia govoria – samy drevny russky iazyk [The Pomor dialect is the oldest Russian language]. Istoriia Severnoy Rusi [The History of Northern Rus], 2013, ukhtoma.ru/chud11.htm (accessed 10 August 2023).

[3] Ilkovič, V. *Rusińskyj Novinař*. Preshov: Rusyn Writers' Union in Slovakia, 2014. 170 p.

[4] Koli ja byv złodijom liêsu, svoja.org/eseji-bijografiji/koli-ja-byv-zlodijom-liesu (accessed 10 August 2023).
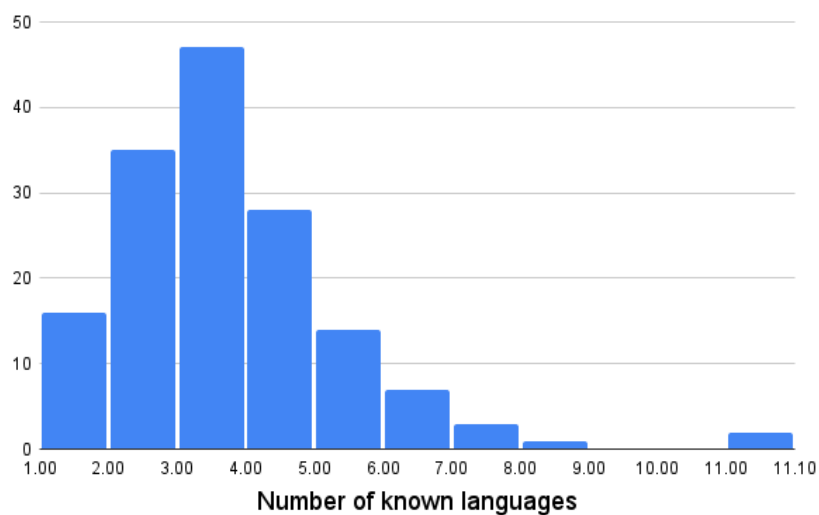
Figure 1. The distribution of participants according to the number of languages they know
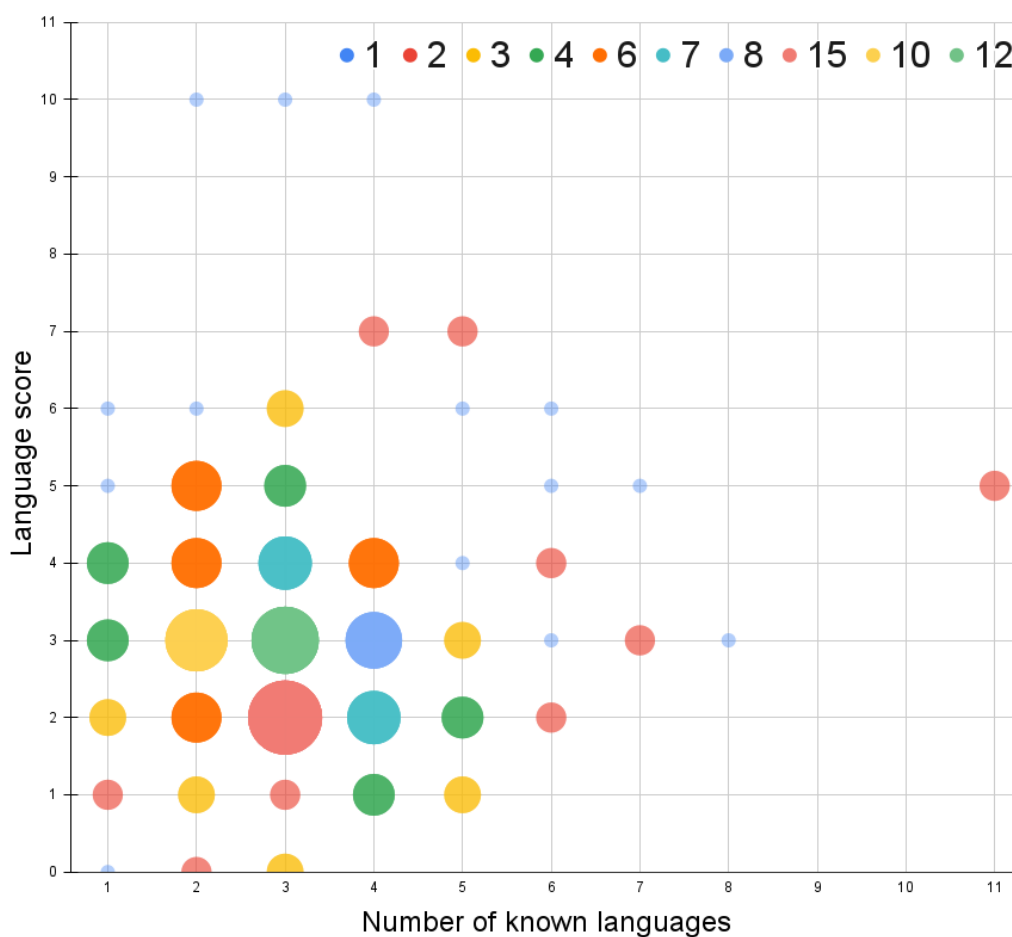

Figure 2. A bubble chart representing the dependence of the language score acquired by a participant vs. the number of known languages. The size of the bubbles represents the number of participants who acquired a given language score while knowing a certain number of languages.

Results and discussion

First, the correlation between the number of languages a person knows and his predisposition to perceive a lect as a language or a dialect was analysed. Task 6 on the list (see the paragraph on methodology) was to identify the lect either as a dialect of Russian or a distinct East Slavic language. The responses were converted into numerical values, with one standing for "language" and zero standing for "dialect", and the "language score" that a participant received was tallied. Thus, if a participant identified seven of the lects as dialects, and three as languages, his "language score" would be three. The author then made a bubble chart with the number of languages that a participant knows on the X-axis, the "language scores" on the Y-axis, and the number of speakers who know a certain number of languages and received a certain "language score" represented by the size of the bubbles (Fig. 2).

Immediately, it becomes apparent that over half of the participants believe there to be three or more distinct languages in the pool of lects. If the initial hypothesis were true, namely, if knowing few languages corresponded to perceiving fewer different languages in a pool of lects, and knowing many languages corresponded to perceiving a greater number of different languages in a pool of lects, one would expect to see a steady upward trend, with the centre of gravity of the bubbles shifting towards the top, as presented in Fig. 3.
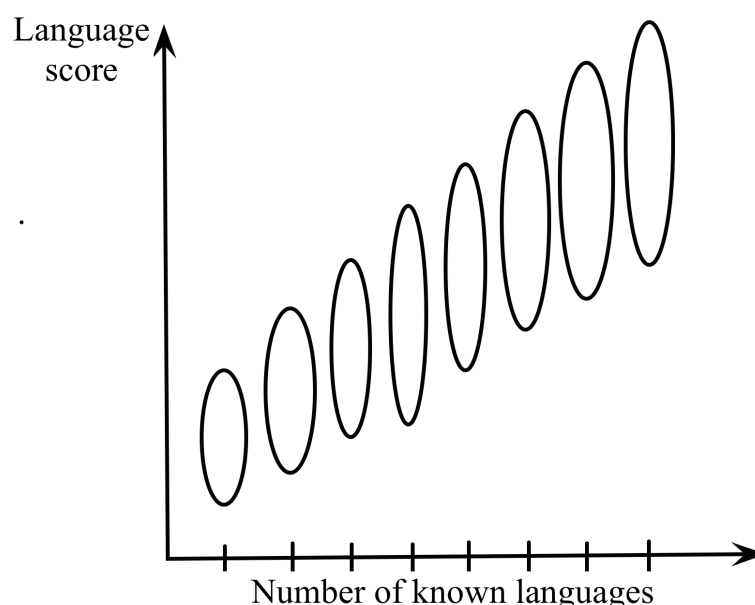

Figure 3. The hypothetical distribution that one should have observed if the initial hypothesis were correct

In reality, however, one sees an uneven distribution with no clear trend. The centre of gravity is always on either two or three "language points" for respondents who know one to five languages. The one interesting feature in support of the author's hypothesis is that among the participants who know over seven languages, none believed there to be fewer than three different languages in the pool of lects. However, given the extremely low number of participants who know more than six languages, such data cannot be considered reliable. In conclusion, it can be stated that the findings demonstrate a weak positive correlation between the number of languages that an individual knows and his predisposition to perceive a lect as a language.

Next, in order to establish the criteria for identifying a lect as a language or a dialect, it was decided to analyse the participants' opinions about the texts presented and look for patterns in their responses. To begin with, the author tallied up the number of "language points" attributed to a language by all of the participants and averaged it across the number of participants, giving the "language score" metric. This

gave him a scale from zero to one that shows how the participants perceive the lect: if the value is closer to one, then it is, without a doubt, a distinct language, according to the participants, but a value that is closer to zero is indicative of a dialect. After carrying out the calculations, the author selected his benchmark lects that would serve as prime examples of what the participants consider to be a language and what they consider to be a dialect. The resulting data is presented numerically in Table 1 and visually in Fig. 4.

	Belgorod	Rusyn	Smolensk1	Samara	Pomor	Smolensk2	Nizhny Novgorod	Podlachian	Yegorievsk	Novgorod
Average	0.24	0.88	0.42	0.08	0.43	0.17	0.33	0.30	0.11	0.25
Confidence interval	0.07	0.05	0.08	0.04	0.08	0.06	0.08	0.07	0.05	0.07

Table 1. The "language scores" of the texts presented to the survey participants.
The confidence intervals were constructed according to Student's *t*-distribution, $p = 0.95$, $n = 153$.

	Belgorod	Rusyn	Smolensk1	Samara	Pomor	Smolensk2	Nizhny Novgorod	Podlachian	Yegorievsk	Novgorod
Average comfort	5.94	3.22	4.36	6.80	4.22	6.28	4.80	6.56	7.36	5.95
Confidence Interval	0.41	0.41	0.46	0.40	0.40	0.43	0.42	0.40	0.36	0.40
Median	6.0	3.0	4.0	7	4	7.0	5	7	8	6
Mode	5.0	0.0	5.0	9	5	9.0	6	9	8	7
Average intelligibility	8.07	4.65	5.76	8.31	5.38	7.76	6.38	8.07	8.94	7.42
Confidence Interval	0.31	0.41	0.43	0.30	0.40	0.37	0.41	0.32	0.23	0.33
Median	9.0	4.0	6.0	9	5	8.0	7	9	9	8
Mode	10.0	5.0	8.0	9	5	10.0	8	10	10	8

Table 2. The statistics for the comfort and intelligibility ratings given to the texts by the survey participants.
The confidence intervals were constructed according to Student's *t*-distribution, $p = 0.95$, $n = 153$.
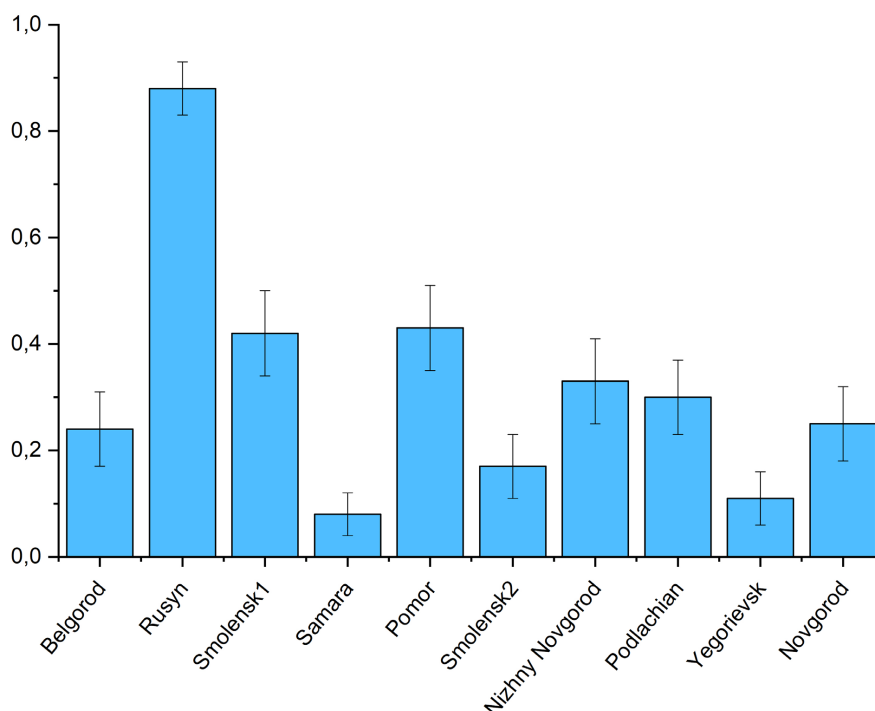

Figure 4. "Language scores" given to the lects by the participants.
Confidence intervals calculated using Student's *t*-distribution, $p = 0.95$, $n = 153$.

The benchmark for a distinct East Slavic language is Rusyn with a "language score" of 0.88 and a 95% confidence interval of 0.05 (according to Student's t-distribution), landing it securely in the "language" category. This is unsurprising given how distant Rusyn is from standard Russian in all aspects: phonetics, lexicon, and syntax. Thus, it can be said that the upper limit for a language criterion lies at around 0.83, which is the average value minus the confidence interval.

The lects that represent the benchmark for dialects of Russian are the dialects of Samara and Ye-gorievsk, with "language scores" of 0.08 and 0.11, respectively, and with 95% confidence intervals of 0.04 and 0.05, respectively (according to the Student's t-distribution). Thus, it can be said that the lower limit for a dialect criterion lies at around 0.12, which is the average for the Samara dialect plus the confidence interval.

Outside of these clear-cut cases, every other lect is more difficult to interpret. Some of the most notable oddities include the following:

- the Pomor "dialect" has a score of 0.43, similar to that of the Smolensk "dialect", with a score of 0.42 (confidence interval 0.08 for both);
- Podlachian has a score of 0.30, similar to that of the Nizhny Novgorod "dialect", with a score of 0.33 (confidence interval 0.07 and 0.08, respectively).

These two results are significant because the Pomor "dialect" is believed by some to be a distinct lan-guage, owing to its highly divergent phonetic system, specific vocabulary and a rather atypical grammar for a Slavic language, which includes clitic definite articles, similar to the Scandinavian languages, Bulgar-ian and Macedonian[5]. These features alone would warrant Pomor the status of a language, so one might be tempted to set the threshold for differentiating a language from a dialect at 0.5 on the "language score". If 0.5 does not fall within the confidence interval of a lect, it is to be considered a dialect, and if 0.5 does fall within the confidence interval of a lect, or the mean is equal to or higher than 0.5, it is to be considered a distinct language. A threshold value of 0.5 is logically reasonable, as it represents the midpoint in terms of speakers' perception, and, additionally, this choice is corroborated by data for the Pomor "dialect". However, since the Smolensk "dialect" has an almost identical score, that would imply that it, too, ought to be considered a separate language, which is a more controversial claim that would require further in-vestigation and elucidation.

On the other hand, the fact that Podlachian scored so low presents another complication for the au-thor's theory. The origin of Podlachian is known and well-documented: it is the East Slavic literary mi-crolanguage that is most closely related to Belarusian, while geographically it is located entirely within the borders of present-day Poland, and its development took place at the turn of the previous century[6]. By all accounts, Podlachian should be as distant from Modern Standard Russian as Prešov Rusyn, which has a similar history and geographic position. Yet, despite every indication that Podlachian should be a language in its own right, its score is remarkably close to that of the Nizhny Novgorod "dialect", both of which are significantly lower than the scores for the Pomor and Smolensk "dialects". Potential causes of this dilemma are:

1. The text that was presented to the participants gave an account of an old man's recollections of nature around his home village, and the vocabulary in this passage is (a) concrete, and (b) cognate with similar words in Modern Standard Russia. It could very well be that this similarity in vocab-ulary and the lack of major innovations in Podlachian phonetics (as compared with, for instance, changes in the mid vowel quality in closed syllables in Prešov Rusyn) allowed for unimpeded com-prehension of the text by speakers of Russian. Meanwhile, a similar effect did not take place in the case of Rusyn, because its phonetics is so strikingly different, so cognates are more difficult to

---

[5] Galanin, A. Pomorskaia govoria – samy drevny russky iazyk [The Pomor dialect is the oldest Russian language]. Istoriia Severnoy Rusi [The History of Northern Rus], 2013, ukhtoma.ru/chud11.htm (accessed 10 August 2023).

[6] Svoja.org (accessed 10 August 2023).

recognise, and the text deals with matters that were more contemplative and abstract in nature. Besides, as East Slavic languages outside of Russia tend to loan abstract terms from Polish, the terms were not cognate with those of Russian and good comprehension was not achieved.

2. Podlachian, and by extension Belarusian, from which it originates, ought to be considered a single language with Russian, albeit with highly divergent dialect groups. As could be seen from the literature review, computational methods do point one in the direction of this hypothesis [9].

If Proposition 1 were true, then one should observe an increase in the language score for Podlachian when a text with a higher content of abstract vocabulary that is not cognate with Russian terms is used. If Proposition 2 is true, then all Podlachian texts will be easily comprehensible to an average speaker of Russian. In both cases more testing with different texts is required.

Additionally, in Questions 2 and 3 in the survey, the participants were asked to rate the intelligibility of the text and their comfort when they imagine speaking to a native speaker of the lect in question. The author averaged the results for every language, built a confidence interval and found the median and the mode. The results are presented in Table 2 and Fig. 5 and 6, respectively.
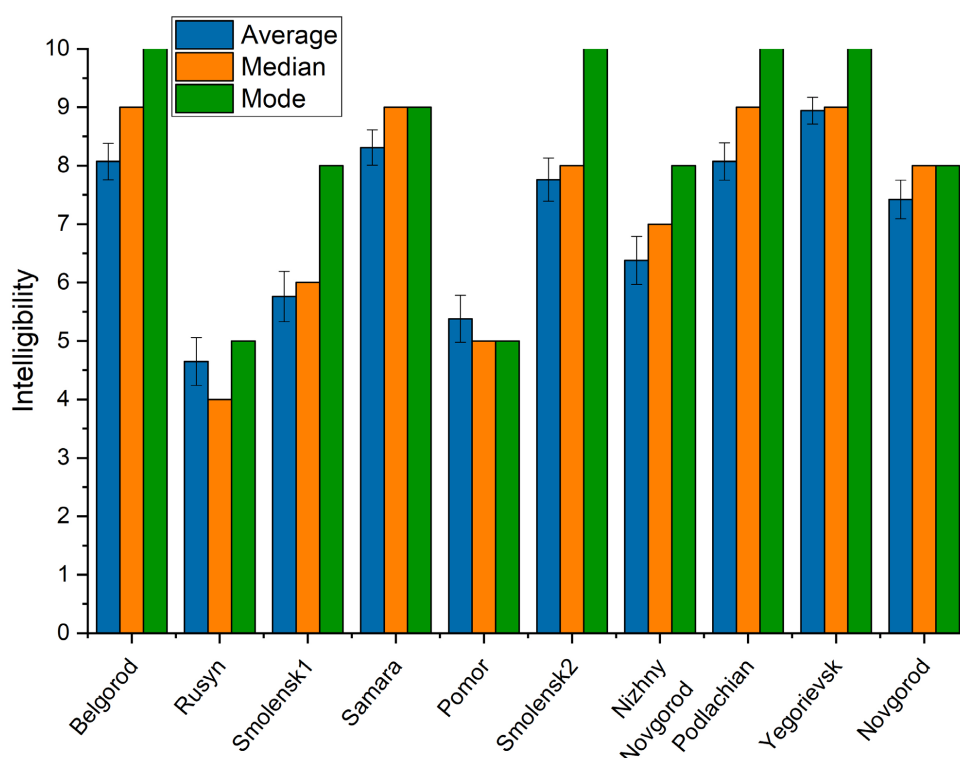

Figure 5. Intelligibility of the lects. Confidence intervals calculated using Student's *t*-distribution, $p = 0.95$, $n = 153$.
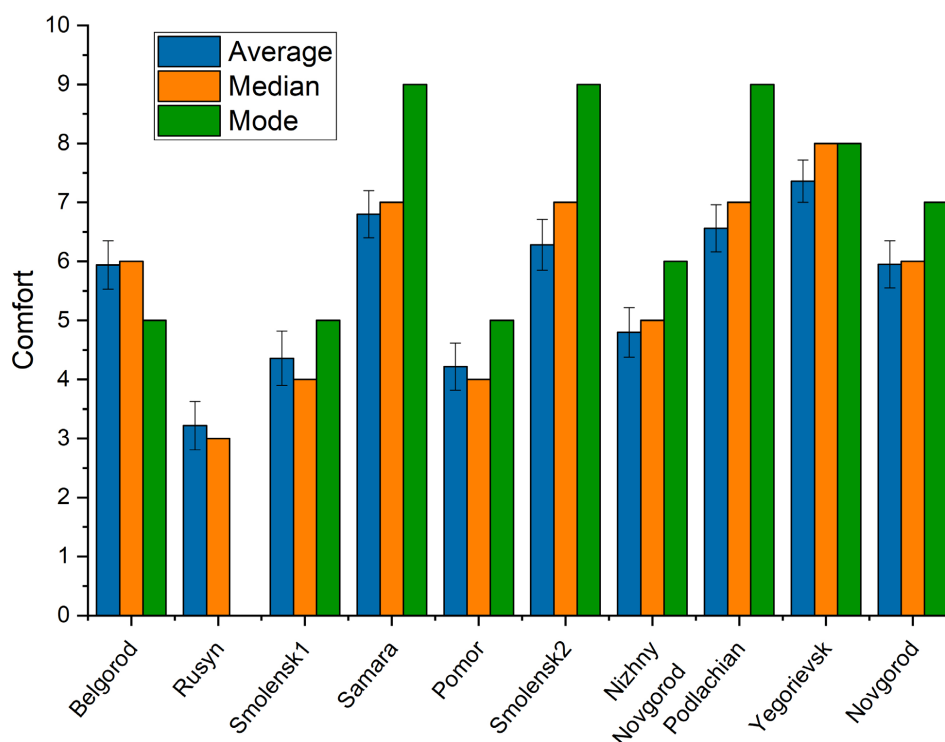
Figure 6. Participants' comfort when dealing with a native speaker of a lect.
Confidence intervals calculated using Student's *t*-distribution, $p = 0.95$, $n = 153$.

The trends noticeable in these data are identical to those observed in the "language scores" dataset. Rusyn stands out as the language with the lowest comfort and intelligibility ratings, with Pomor coming in at number two. The other lects have similar ratings, so these data will not be analysed further.

Factors influencing dialect perception

In Questions 4 and 5 of the survey, the participants were asked to name the features that, in their opinion, were the most divergent from and the most similar to Modern Standard Russian. In each case they were allowed to choose up to three of the following options: phonetics, vocabulary, and grammar. It was decided to analyze the trends in the data obtained to establish whether certain features take precedence over others in determining a lect as a language or a dialect.

The number of times a given feature was mentioned for every lect was calculated, both as a similarity and as a difference. These responses were then sorted into two categories for every lect: the responses given when the lect was identified as a language, and the responses given when the lect was identified as a dialect. The author summed the results for every category over every lect and divided the results by the number of participants to get a number between zero and ten, presented in Fig. 7 (which shows the number of times an average speaker named a given feature a similarity or a difference after identifying the lect as a language) and Fig. 8 (which shows the number of times an average speaker named a given feature a similarity or a difference after identifying the lect as a dialect). These numbers allow one to compare the importance that a given characteristic has when a participant has to identify a lect as a language or a dialect.
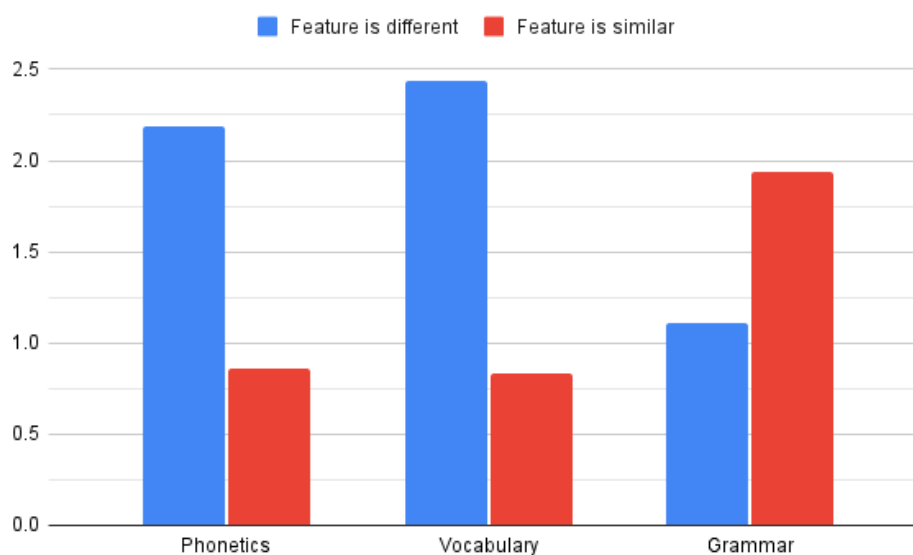
Figure 7. The number of times that a feature was deemed different or similar when a lect was identified as a language, averaged across the participants
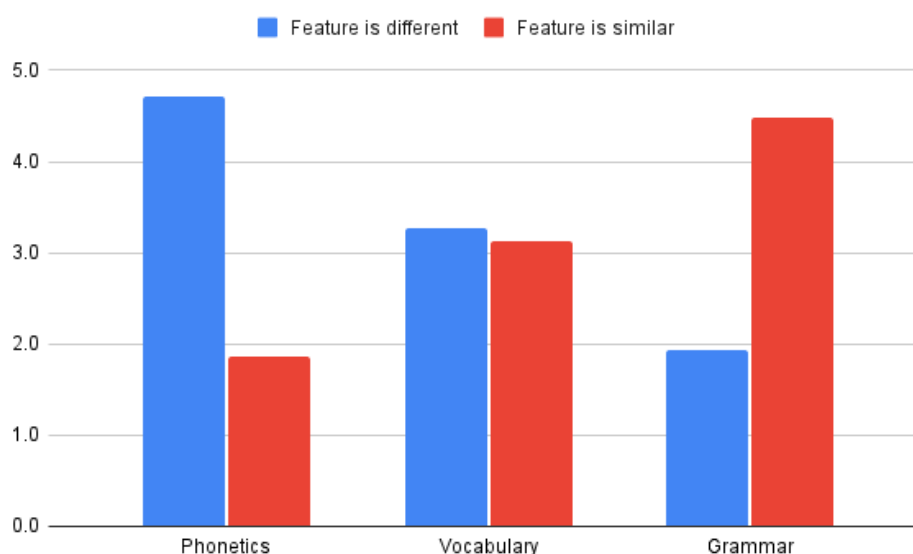

Figure 8. The number of times that a feature was deemed different or similar when a lect was identified as a dialect, averaged across the participants

In the case of a participant identifying a lect as a language, one observes that, on average, speakers would consider a lect's phonetics significantly different from that of Modern Standard Russian 2.2 times; a lect's vocabulary, 2.4 times; and grammar, 1.1 times. Meanwhile, these same features were considered to be close to those of Modern Standard Russian 0.9, 0.8, and 1.9 times, respectively. Thus, it can be seen that, when identifying a lect as a language, speakers tend to consider the phonetics and the vocabulary of the lect to be the key deciding factors. Meanwhile, the participants consider the grammar of the lects to be fairly close to that of Modern Standard Russian, which is reasonable, since all of the lects belong to the East Slavic branch, and thus all share most of their grammatical features.

In the case of a participant identifying a lect as a dialect, one observes that, on average, speakers would consider a lect's phonetics significantly different from that of Modern Standard Russian 4.7 times; a lect's vocabulary, 3.3 times; and grammar, 1.9 times. Meanwhile, these same features were considered to be close to those of Modern Standard Russian 1.9, 3.1, and 4.5 times, respectively. In stark contrast to the case of a lect being identified as a language, here the participants named vocabulary different and similar an almost identical number of times, indicating that in the respondents' view, vocabulary does not play a role in recognising a dialect. Meanwhile, phonetics does appear to be of importance for identifying a dialect, as it was designated as different almost 2.5 times more often than as similar. Grammar, on the other hand, has a similar distribution as in the case of a lect being identified as a language, which makes even more sense in this case as the dialects of a single language cannot be substantially different.

Thus, it can be concluded that in the participants' view, a dialect differs from the standard language only in its phonetics, while a closely related language differs from their mother tongue both in phonetics and in vocabulary, but not in grammar, due to them being closely related.

The author then compared the distributions of similar and differing features for three key lects: Rusyn, the Samara "dialect" and the Pomor "dialect". These three were chosen because they represent a clearly distinct language, a clear dialect, and a controversial "middle-of-the-road" lect. By considering the distributions of similar and differing features, one can find additional ways of comparing the different lects. The results are presented in Fig. 9 (displaying the differences) and Fig. 10 (displaying the similarities).
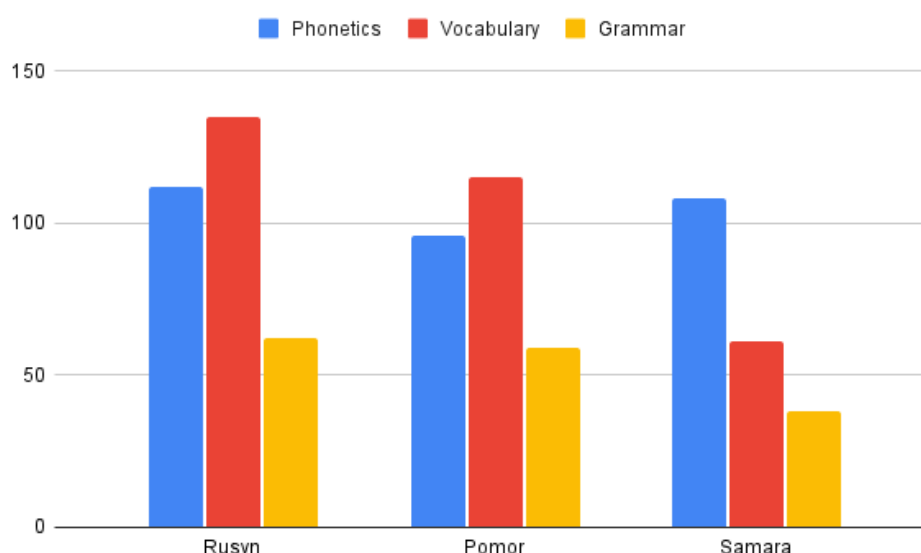

Figure 9. The number of times that a feature was deemed to be a difference for a given lect. The remaining lects showed a distribution similar to that of the Samara "dialect".

When comparing Rusyn and the Samara "dialect", one sees that, just as in the global case discussed above, both vocabulary and phonetics are more frequently deemed different in Rusyn than in the Samara "dialect", where only the phonetics is seen as substantially different. Grammar is not deemed significantly different in either lect. Meanwhile, when looking at the similarities between the lects and Modern Standard Russian, the distribution is completely flipped: for Rusyn, the phonetics and the vocabulary are rarely seen as similar, with grammar being by far the most similar aspect; and for the Samara "dialect", vocabulary is also seen as fairly similar to that of Modern Standard Russian. This more detailed look supports the conclusion that was made in the preceding paragraph. Additionally, it can be observed that Pomor displays a distribution identical to that of Rusyn, which, in addition to all of the other data, makes one conclude that it is far closer to a distinct language than to a dialect.
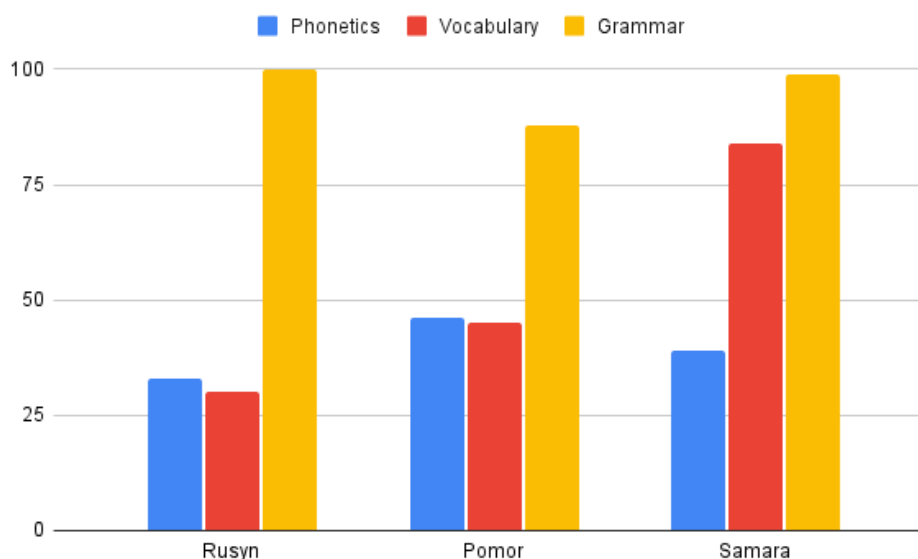
Figure 10. The number of times that a feature was deemed to be a similarity for a given lect. The remaining lects showed a distribution similar to that of the Samara "dialect".

Conclusions

Having analysed the acquired data, the author was unable to confirm his initial hypothesis about polyglottery (represented by the number of languages that a participant knows) correlating with an increased propensity for perceiving closely related lects as distinct languages rather than dialects. Only a very weak correlation was observed, as the number of people in the survey who know more than six languages is too low to make definite claims about the statistical significance of the results. Moreover, an unexpected result was found in that Podlachian, a distinct East Slavic literary microlanguage with a known recorded history, rated higher in comprehension than some dialects of Russian.

What the author did find, however, was the influence that certain linguistic features have on the perception of the lects: the respondents find phonetics to be the sole deciding factor for designating a lect as a dialect, while the combination of differing phonetics and vocabulary are the key deciding factors for recognising a lect as a language.

All this gives one pause and requires further investigation with a greater number of participants and different texts.

———————————

© Ya. Aleshkevich-Suslov, 2023

References

1. Blinova, O., et al. *Praktikum po Russkoy Dialektologii* [Russian Dialectology Practicum]. Tomsk: Tomsk University Press, 1993. 146 p.
2. Clopper, C.G. Perception of dialect variation. *The Handbook of Speech Perception* (ed. by J. Pardoetal). 2nd ed. Oxford: John Wiley & Sons, 2021. P. 334–364.
3. *Dialektologicheskie Materialy, Sobrannye V. I. Trostianskim, I. S. Grishkinym i dr.* [Dialectological Materials, Collected by V. I. Trostiansky, I. S. Grishkin et al.] Petrograd: The Imperial Academy of Sciences Press, 1916. 158 p.
4. Gooskens, C. Dialect intelligibility. *The Handbook of Dialectology* (ed. by C. Boberg et al.). Oxford: John Wiley & Sons, 2018. P. 204–218.

5. Hyltenstam, K. Language aptitude and language awareness: Polyglot perspectives. *Annual Review of Applied Linguistics,* 2021, vol. 41. P. 55–75. DOI: 10.1017/S0267190521000027

6. Kazakov, G. Polyglottery as a science. *International Conference on Multilingual Proficiency: Language, Polyglossia and Polyglottery* (W. Finke & L. Ashley eds.). New York: The American Society of Geolinguistics, 2016. P. 133–137.

7. Lee, R.R. Dialect perception: A critical review and re-evaluation. *Quarterly Journal of Speech,* 1971, vol. 57, no. 4. P. 410–417. DOI: 10.1080/00335637109383086

8. Preston, D.R. Perceptual dialectology. *The Handbook of Dialectology* (ed. by C. Boberg et al.). Oxford: John Wiley & Sons, 2018. P. 177–203.

9. Wichmann, S. How to distinguish languages and dialects. *Computational Linguistics,* 2019, vol. 45, no. 4. P. 823–831. DOI: 10.1162/COLI_a_00366

10. Williams, G. Intelligibility and language boundaries among the Cangin peoples of Senegal. *Journal of Language Survey Reports,* 2021, no. 54. P. 1–23.

Список литературы

1. Блинова О.И. Практикум по русской диалектологии / О. И. Блинова, О. И. Гордеева, Л. Г. Гынгазова. Томск: Изд-во Томского ун-та, 1993. 146 с.

2. Clopper C.G. Perception of dialect variation // The Handbook of Speech Perception / ed. by J. Pardoetal. 2nd ed. Oxford: John Wiley & Sons, 2021. P. 334–364.

3. Диалектологические материалы, собранные В. И. Тростянским, И. С. Гришкиным и др. Петроград: Тип. Императорской академии наук, 1916. 158 с.

4. Gooskens C. Dialect intelligibility // The Handbook of Dialectology / ed. by C. Boberg et al. Oxford: John Wiley & Sons, 2018. P. 204–218.

5. Hyltenstam K. Language aptitude and language awareness: Polyglot perspectives // Annual Review of Applied Linguistics, 2021, vol. 41. P. 55–75. DOI: 10.1017/S0267190521000027

6. Kazakov G. Polyglottery as a science // International Conference on Multilingual Proficiency: Language, Polyglossia and Polyglottery / W. Finke & L. Ashley eds. New York: The American Society of Geolinguistics, 2016. P. 133–137.

7. Lee R.R. Dialect perception: A critical review and re-evaluation // Quarterly Journal of Speech, 1971, vol. 57, no. 4. P. 410–417. DOI: 10.1080/00335637109383086

8. Preston D.R. Perceptual dialectology // The Handbook of Dialectology / ed. by C. Boberg et al. Oxford: John Wiley & Sons, 2018. P. 177–203.

9. Wichmann S. How to distinguish languages and dialects // Computational Linguistics, 2019, vol. 45, no. 4. P. 823–831. DOI: 10.1162/COLI_a_00366

10. Williams G. Intelligibility and language boundaries among the Cangin peoples of Senegal // Journal of Language Survey Reports, 2021, no. 54. P. 1–23.

About the author:

Yan Aleshkevich-Suslov is a graduate of the Faculty of Chemistry, Moscow State University, and a teacher at the School of the Centre for Pedagogical Mastery in Moscow. Research interests: chemical thermodynamics, sustainable development, green chemistry, language learning, comparative linguistics, polyglottery, natural theology. E-mail: aleshkevichyan@yandex.ru.

Сведения об авторе:

Ян Вячеславович Алешкевич-Суслов – выпускник химического факультета МГУ им. М. В. Ломоносова и преподаватель школы Центра педагогического мастерства в Москве. Исследовательские интересы: химическая термодинамика, устойчивое развитие, зелёная химия, изучение языков, сравнительная лингвистика, полиглотия, естественная теология. E-mail: aleshkevichyan@yandex.ru

* * *