NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Ilya Yu. Chechuro

QUANTITATIVE ANALYSIS OF PITCH MOVEMENT IN SPONTANEOUS SPEECH OF REGIONAL RUSSIAN NATIVE SPEAKERS

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: LINGUISTICS
WP BRP 66/LNG/2018

SERIES: LINGUISTICS

Ilya Yu. Chechuro[1]

QUANTITATIVE ANALYSIS OF PITCH MOVEMENT IN SPONTANEOUS SPEECH OF REGIONAL RUSSIAN NATIVE SPEAKERS[2]

The paper deals with the intonation of native speakers of several regional variants of Russian. I analyze how simple characteristics of pitch movement are connected to biological sex, age, place of living, and type of spontaneous text. According to my primary hypothesis, the distribution of pitch in “low”, “middle” and “high” parts of the pitch range and the number of movements from one range into other ones can predict the aforementioned characteristics of the recordings as well as reveal the individual characteristics of speakers (Vol’skaya 2014, Féry 2017). I demonstrate that in my model the only significant parameter is the biological sex of the speakers, which means that this distribution is unlikely to be determined by the regional differences. The study has been conducted as a part of the project on statistical analysis of regional variants of Standard Russian and minority languages of Russia.

JEL Classification: Z

Keywords: Russian, regional variants of Russian, Intonation, quantitative analysis, statistical models.

[1] National Research University Higher School of Economics. School of Linguistics. Assistant. / Linguistic Convergence Laboratory. Junior research fellow. E-mail: ilyachechuro@gmail.com

[2] The article was prepared within the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE) and supported within the framework of a subsidy by the Russian Academic Excellence Project '5-100'.

1. Introduction

The study of intonation has by now become a relatively well-established area with different schools and approaches. The two main approaches to intonation can be characterized as *qualitative* and *quantitative*. The qualitative studies such as (Odé 1989, Vol'skaya 2014, Yanko 2017) usually suggest classifications of intonation patterns and their semantic interpretations. There is also a number of prescriptive studies related to language teaching and language description (Bryzgunova 1963).

One of the first studies of regional intonation in Standard Russian has been conducted by (Karinskiy 1929), who discussed the features of the intelligentsia speech in Vyatka (Kirov) including the particularities of the intonation. Though this study was more of a description of Karinskiy's own impression of the intelligentsia speech and did not imply any unbiased analysis, it was also the first one where the existence of the regional variants of standard Russian was postulated. This idea is further developed by Scherba (1957: 56) and Panov (1967: 294) who emphasize the idea that Standard Russian dialects do exist and that they should not be viewed as a transition stage on the path from dialectal to standard pronunciation.

A relatively wide range of papers deal with the remaining (Rus. *остаточные*) features of dialectal speech in regional Standard Russian (Parikova 1966, Bondarko and Verbitskaya 1987, Erofeeva 1979, Almukhamedova and Kulsharipova 1980, Paufoshima 1983, and Erofeeva 1997). Though these papers mostly deal with segmental characteristics of regional speech, they also discuss the methods of oscillographic and spectral analyses.

One of the important findings of these studies is the existence of prosodic differences between Standard Russian and its regional variants, namely the lower salience of the 1st pre-stressed syllable and even the absence of the two-syllable prosodic core of a word form (Vysotsky 1973, Avanesov 1984, Almukhamedova and Kulsharipova 1980, Paufoshima 1983, Daniel et al. 2010).

A recent paper by Grammatchikova et al. (Ms) deals with the rhythmical structure of word forms and the realization of tonal accent in regional variants of Russian. The authors use an experimental approach and apply their analysis to comparable data from different region and demonstrate that the all the phases of the pitch change due to the phrasal accent happen earlier in

time and closer to the left boundary of a word in Standard Russian than in several other Regional Russian variants.

Despite the abundance of qualitative research dealing with the intonation in regional variants of Russian, the quantitative studies are mostly limited to the Standard Russian spoken in Moscow or in St. Petersburg. The scholars engaged in the quantitative studies of intonation usually attempt at building corpora (cf. One Speaker's Day (Stepanova et al. 2008), CORUSS (Kachkovskaia et al. 2016)) and investigating the numeric parameters of intonation, such as pitch frequency, intensity and others (Šimko et al. 2017).

In this study, I perform a quantitative analysis of intonation in several regional variants of Standard Russian. Using the recordings made in four different regions of Russia (Krasnoyarsk, Moscow, Nakhodka and Novosibirsk) I analyze the pitch movement in spontaneous speech of the native speakers of regional Standard Russian. For each speaker, I recorded three samples of spontaneous speech: an interview, a dialogue and a retell of the pear movie (Chafe 1980). According to my primary hypothesis, the distribution of pitch in "low", "middle" and "high" parts of the pitch range and the number of movements from one sub-range into the other ones can predict the aforementioned characteristics of the recordings as well as reveal the individual characteristics of speakers (Vol'skaya 2014, Féry 2017). Using linear mixed effect modelling, I show that the significant factor for the amount of pitch movement is the biological sex of the speakers, while the factors of place and text type are not significant. According to my data, men tend to have most of their pitch values within the "Low" part of the range, while women use other parts of the range more often. Furthermore, in male speech there are less transitions between the pitch levels than in female speech.

These results indicate that there is a significant difference between men and women in the pitch use. These differences can be explained in two different ways: (a) male speakers use pitch movement more rarely than female speakers and (b) the pitch movement in males has a lower amplitude in males but it is not necessarily less frequent. Both interpretations, however, suggest that there is a major difference between male and female pitch use and only differ in the nature of this difference. Furthermore, according to my data the amount of pitch use does not seem to be responsible for the regional differences.

The remainder of this paper is organized as follows. Part 2 discusses the participants and the experimental conditions. Part 3 discusses the data sampling. Part 4 deals with the preliminary

analysis of the pitch values. Part 5 contains the description of the statistical analysis of the data. Part 6 concludes the paper. Part 7 outlines the further research. Part 8 is the bibliography.

2. Participants and Experimental Conditions

The spontaneous dialogues were recorded in the "fieldtrip" conditions in a quiet room using a recorder that supports *.WAV* format with no compression. The recordings made in Moscow, including those with the regional respondents, were made with a professional recorder and individual headset microphones for each speaker.

All participants were monolingual native speakers of Russian born in Krasnoyarsk, Novosibirsk, Nakhodka, and Moscow. Krasnoyarsk and Novosibirsk represent Siberia, Moscow — Standard Russian and Nakhodka — Far East (the city population of which usually originates from different regions of the ex-USSR and is highly mixed). At the moment of the recordings, all the participants lived in their home regions or have recently moved to Moscow to study at the university (1st year students in the beginning of the 1st semester). All regional participants were divided into two age groups: from 25 to 40 years old vs. 45 years old and older. This division was made in order to balance the sample; in the analysis presented in this paper age was used as a numeric and not as a categorical variable. In each age group, there were two male and two female participants. The speakers from Moscow were represented by two females from the lower age group.

Each recording has been taken from two participants. In all pairs, the interlocutors knew each other relatively well (they were classmates, friends or relatives) and belonged to the same age and social group.

The experiment began with setting up the recording devices and instructing the participants. This stage took from 5 to 10 minutes. During this time, the participants could talk to each other freely and simultaneously get used to the recording equipment and the experimental environment.

2.1. Tasks for the Participants

After the recording equipment has been set up, the participants were orally instructed with the tasks for the experiment. There were three types of tasks: interview, experiment with a map and storytelling based on a pear movie (Chafe 1980). In the first task, the participants had to tell a small story about their life (e.g. parents and family, school, favourite teachers, hometown, etc.). The story

had to last from 8 to 10 minutes. If it was necessary, questions were asked to the participant in order to continue the elicitation.

The second task was an experiment with a map based on (Usacheva Ms.). Two participants, the instructor and the follower, were given a map of the Moscow Zoo (Fig. 1) printed on an A2 sheet and a set of objects (coins, pencils, dices, etc.) to place on the map. The sets were the same for both participants with the exception for one or two elements. The differences in the sets were designed specifically to provoke a mismatch between the instructor and the follower and thus to force them to use exclamations and questions. The participants were seated in different rooms, so that they could not hear or see each other.

Fig. 1. The Map of the Moscow Zoo

During the experiment, the speakers communicated using mobile phones. The experimenter placed objects on the Zoo map in front of the instructor. The instructor had to explain the positions of these objects to the follower so that the follower could repeat it on his map. The follower was assisted by a second experimenter who controlled how the objects were placed and helped the follower if it was necessary. The active phase of the experiment was preceded by a training phase,

during which the experimenter could make sure that the participants understand the instructions correctly and that there are no technical failures. This part of the experiment took from 20 to 25 minutes.

The third part of the experiment implied retelling the Pear Movie (Chafe 1980). The recording lasted from 2 to 5 minutes.

3. Data Sampling and Annotation

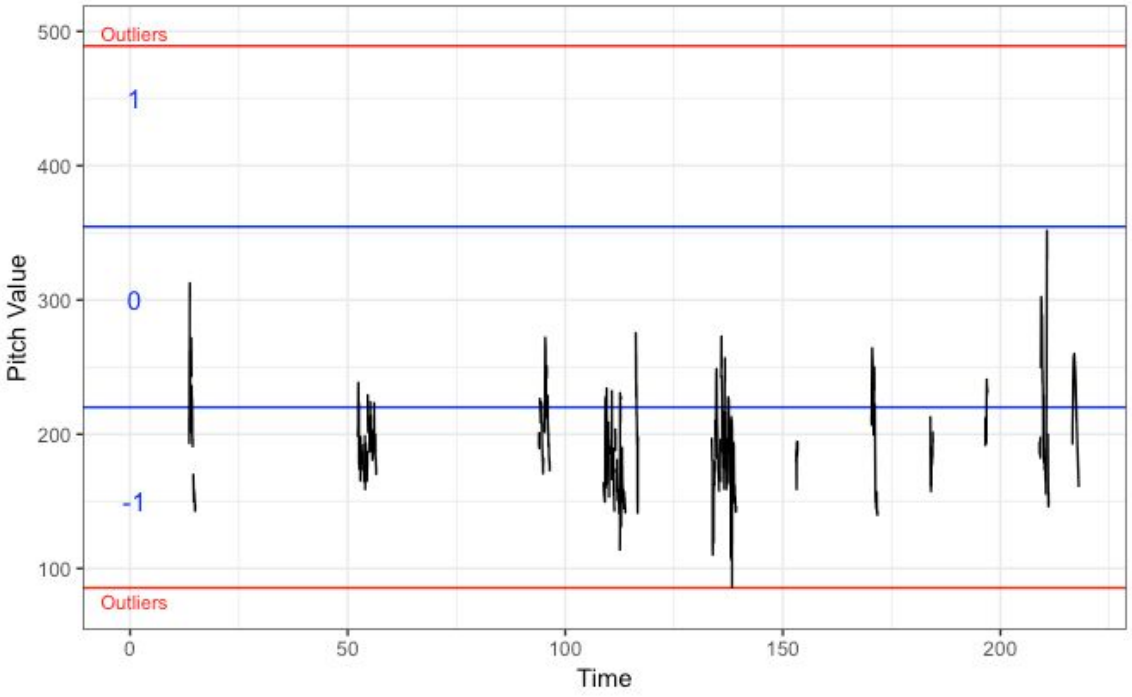
Each speaker in the dataset was represented by 40 randomly selected utterances. The utterances were extracted from each recording type using the following proportion: 15 recordings from the interview, 15 — from the experiment, and 10 from the pear story. There length of the recordings was not normalized. The pitch values were extracted from each recording with a 10 ms step. The pitch values were extracted with the standard functions of Praat. The pitch range (maximal and minimal pitch values) was defined for each speaker separately.

The data have been annotated in Praat. The first TextGrid tier contained the boundaries of the speech units defined by pauses on the oscillogram. Other tiers contained the markup for questions, exclamations, abrupt utterances, non-language sounds (hesitation pauses, breathes, etc.). The parts of the recordings that contained sounds other than the participant's speech (experimenters' instructions, random noises) or technical problems such as distortion or low volume were marked on a separate tier. These parts of the recordings were not used in the analysis.

4. Analysis of the Pitch Value Distribution

The analysis of the data presented in this section partly adopts the approach introduced in (Šimko et al. 2017). Smoothing and wavelet transformations were omitted in my analysis. For each of the speakers, the pitch range was defined as the difference between the minimal and the maximal pitch values in all 40 recordings, with the exclusion of 5% of the observations: 2,5% with the minimal values and 2,5% with the maximal values (Fig. 2). This type of range narrowing lowers the probability of including octave jumps and other artefacts into the analysis. Then, the pitch values were normalized by the *z-score*. The remaining range (95% observations) was divided into three equal parts that were coded, respectively, with -1 (Low), 0 (Medium) и 1 (High), which correspond to the commonly used division of the pitch range into Low, Medium and High (Keijsper 2003, Odé 1989).

Fig. 2. Pitch sub-ranges in a recording sample



At the second step of annotation the transitions between the -1, 0 and 1 levels were encoded. The data were annotated as follows: if the points N and N+1(taken with a 10 ms interval) are in the same pitch level, I interpret this as no-change in pitch shape and code it as 0. If the points are in different sub-ranges, the transition is coded as the difference between the levels: -2 (High to Low), -1 (Medium to Low, High to Medium), +1 (Low to Medium, Medium to High) или +2 (Low to High). This annotation was designed in order to distinguish substantial pitch movements from its minor fluctuations within a single sub-range.

The two types of the annotation were used to compose four datasets: the distribution of observations by the sub-ranges, the transitions between the sub-ranges, and the number of observations in each sub-range and the transitions of each type.

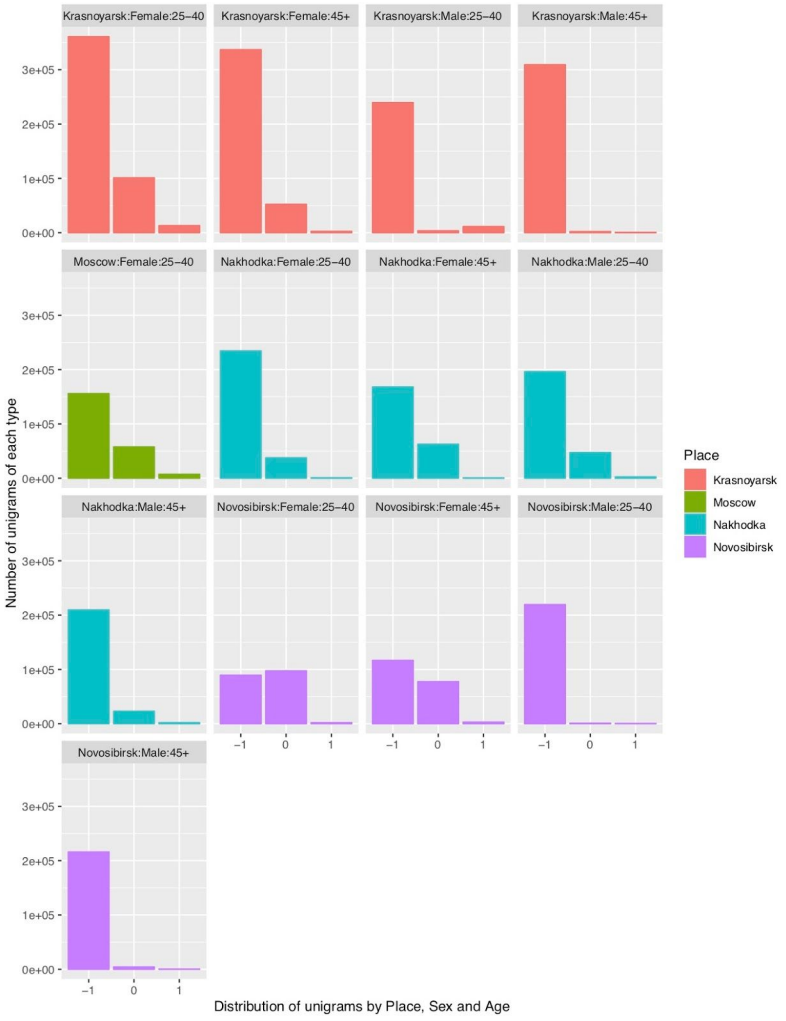
Each line of the dataset corresponded to one pitch value and contained the information about the speaker's name, their place of living, biological sex, age, text type, sentence ID from 1 to 40, time on the recording the observation corresponds to, the sub-range value -1, 0 or 1 (further called *unigram*) or the transition value -2, -1, 0, 1, 2 (further called *bigram*). Upon this dataset, I created a new one, where each line corresponded to a sentence and the rows contained the meta information about the text and the number of unigrams or bigrams of each type in this sentence.

Due to the size of the datasets of the first type (the size of each dataset in the *.csv* format was over 300 megabytes) and the limited resources of the personal computer used for statistical modelling, the data of the first dataset were not used in the current study. Nevertheless, I plan to use these data for statistical modelling using specifically designed systems with a better performance. The analysis presented in this paper was conducted using only the second type of datasets.

5. Data Analysis

The preliminary analysis of the data was conducted using histograms of the unigrams and bigrams values per speaker[3]. Figure 3 illustrates the distribution of unigrams per speaker, the histograms are colored by the region:

Fig. 3. The Distribution of the z-scored Unigram Values by Place, Sex and Age

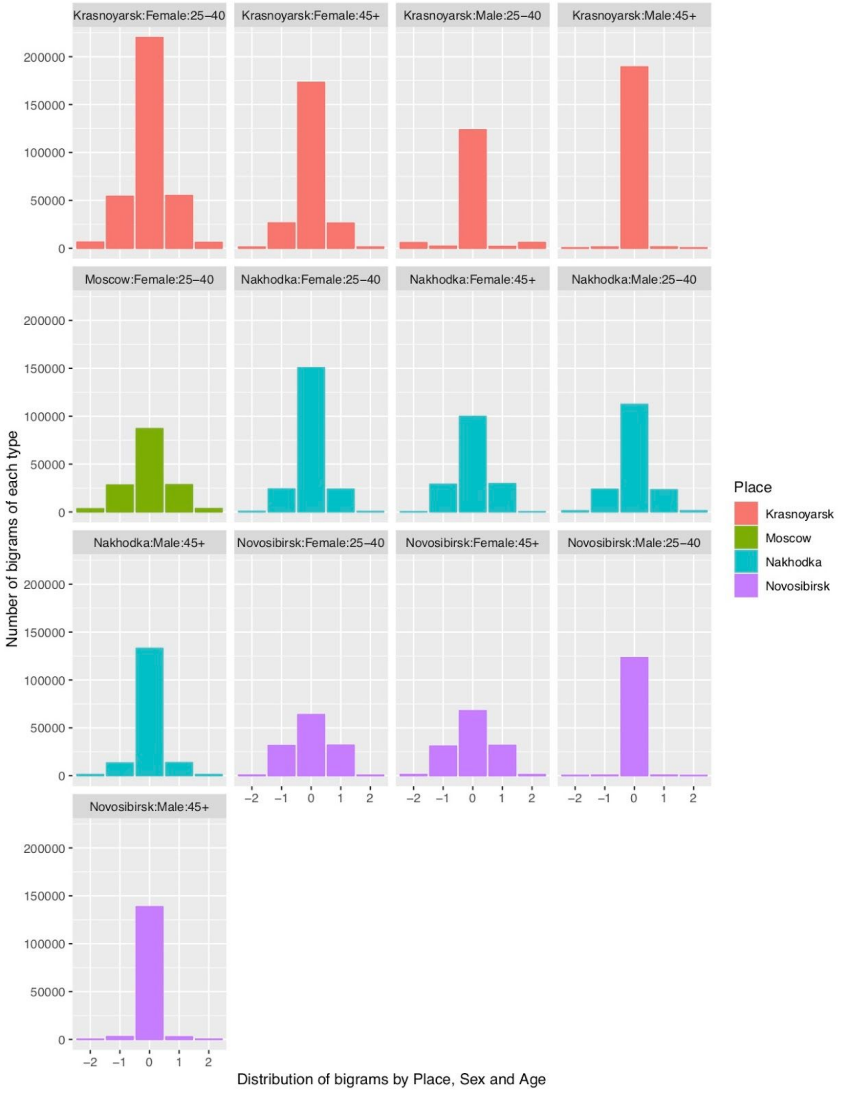


[3] The analysis presented in this part was conducted using the R programming language and the *lme4*, *effects, Hmisc, ggplot2* и *MASS* extension packages.

The shape of the histograms in Figure 2 suggests that the difference between male and female respondents may play a significant role in Novosibirsk and Krasnoyarsk with women having a greater proportion of "0" unigrams, while in Nakhodka this difference is less pronounced and both sexes are similar. The histograms also indicate that the number of "1" unigrams may not be of much significance and the opposition can be viewed as "-1" vs. "not -1" values.

Figure 4 illustrates the distribution of bigrams by speaker. The shape of the histograms suggests that the main difference between male and female respondents is in how often the "-1" and "1" bigrams occur in their recordings. Again, in Nakhodka this difference is less pronounced than in other regions. Apart from three speakers, the number of -2 and 2 bigrams is imperceptible and the main opposition appears to exist between 0 vs. -1 and 1 bigrams. The data can thus be coded as "0" vs. "not 0" bigrams indicating the presence and the absence of pitch movement.

Fig. 4. The Distribution of the z-scored Bigram Values by Place, Sex and Age



In the following analysis of the data I conducted a linear mixed effect modeling to explore the effects of biological sex, age, place of origin and type of text (dialogue vs. monologue) on the distribution of unigrams and bigrams. I fitted the following models. The first model predicted the proportion of "-1" to "not -1" unigrams on the basis of biological sex, age, place of origin of the speakers, the type of the text and the speaker identity as a random effect *lmer(X1Prop ~ Sex*Age + Place + TextType + (1 | Speaker_ID))*. The second model predicted the same value on the basis of sex, age and the type of the text with the place of origin as a random effect *lmer(X1Prop ~ Sex*Age + + TextType + (1 | Place))*. Then, I fitted two models with the same predictors for the proportion of "0" bigrams to the "not 0" bigrams: *lmer(X0Prop ~ Sex*Age + Place + TextType + (1 | Speaker_ID))* and *lmer(X0Prop ~ Sex*Age + TextType + (1 | Place))*.

The predicted value in the first was the proportion of "-1" unigrams to the "not -1" values. The controlled variables were biological sex, age, place of origin and type of text (dialogue vs. monologue) and the speaker ID as a random effect. The stepwise regression model selection with backward elimination has shown that the only significant variable is sex (p-value = 0.014) and the text type is on the edge of significance (p-value 0.06) with random intercepts by speakers. The effect of the two variables is provided in Fig. 5.

Figure 5 shows that the proportion of "-1" unigrams is significantly lower in females than in males. Similarly, the same proportion to a smaller degree is observed with respect to the text type. The factors of place and age turned out to be insignificant.

The second model was bigram-based and predicted the proportion of "0" bigrams to the "not 0" bigrams. The set of predictors was the same as in the first model. The backward model selection has shown that the only significant predictors is biological sex. Figure 6 illustrates that the amount of pitch movement in male participants is significantly lower than that in female participants.

Fig. 5. The effect of biological sex on the proportion of “-1” unigrams to “not -1” unigrams

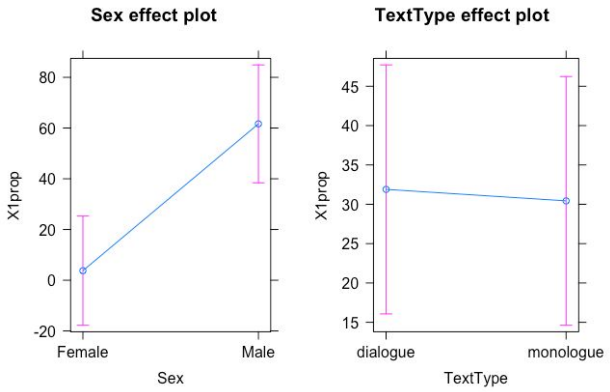
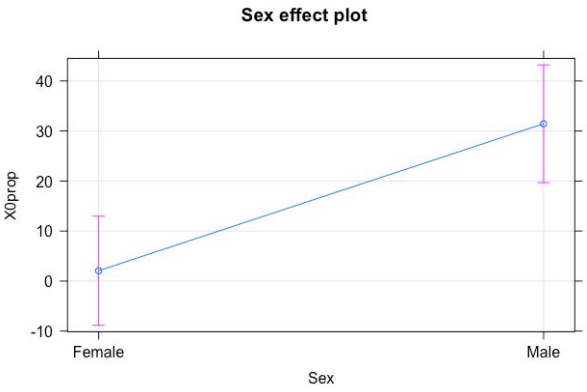


Fig. 6. The effects of biological sex and age on the proportion of “0” and “not 0” bigrams



The unigram-based model I fitted next was similar to the one described above with the only change being made to the random effect structure: instead of fitting random intercepts for speakers, I used random intercepts for different places. The backward stepwise regression model validation has shown that the significant predictors are sex and age with the age-related change in pitch use being significant in men an insignificant in women.

Similar changes were made to the bigram-based model. The significant effects turned out to be the same as in the unigram-based model. Figure 7 and Figure 8 illustrate the effect of age with respect to biological sex. The effect of age is significant in men but not significant in women, which means that as the age increases men start to use pitch movement more. The models thus suggest that though there is no global effect of age it does exist within each city separately.

Fig. 7. Effect of age by sex on the proportion of "-1" unigrams.

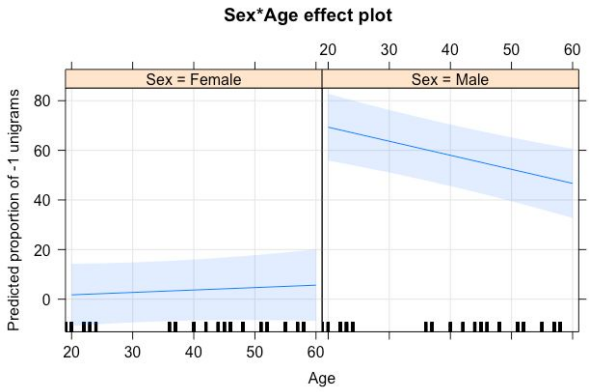
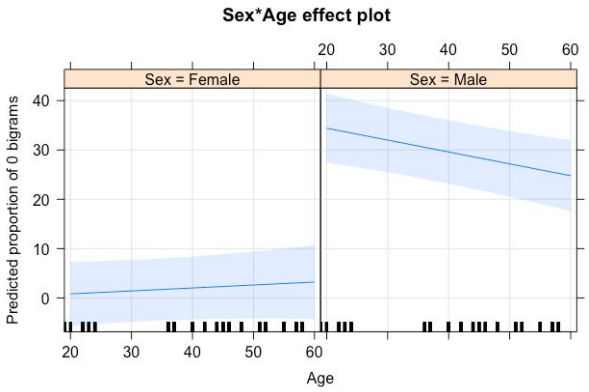


Fig. 8. Effect of age by sex on the proportion of "0" bigrams.



The last pair of models that I have fitted predicted the proportion of -1 unigrams and 0 bigrams on the basis of the interaction between sex, age and place: *lm(formula = X0prop ~ Sex:Age:Place)* and *lm(formula = X1prop ~ Sex:Age:Place)*. The data for Moscow were removed from the dataframe since they only correspond to one age group and one biological sex.

Both models suggest that there is a significant difference between men and women in all cities except for Nakhodka (p-value 0.252 for unigrams and 0.374 for bigrams). Figure 9 and Figure 10 illustrate the effect of age and sex on the proportion of -1 unigrams and 0 bigrams.

Fig. 9. Effect of age and sex by region on the proportion of “-1” unigrams

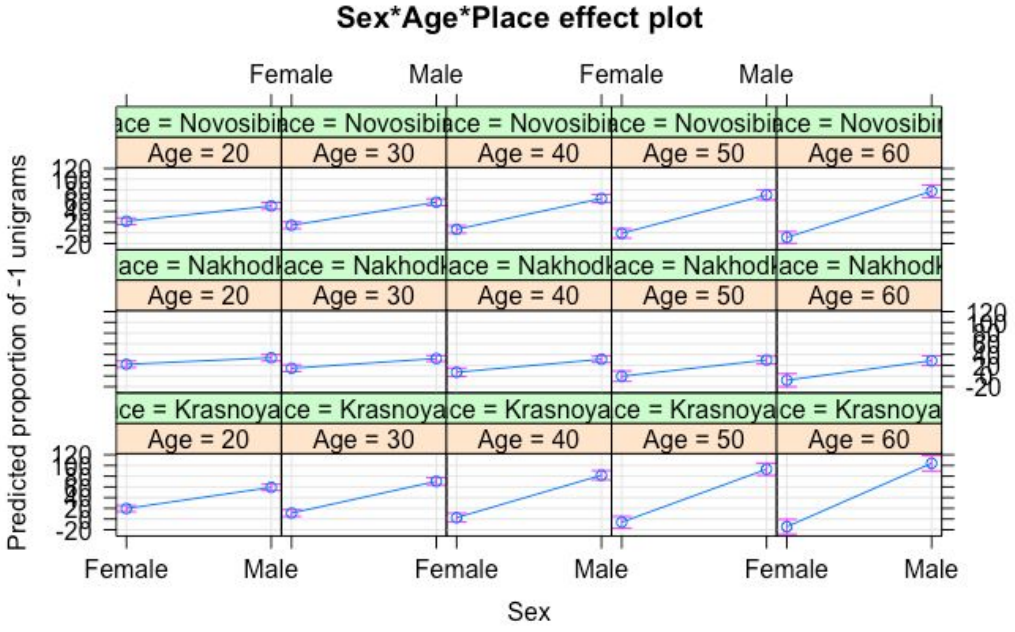
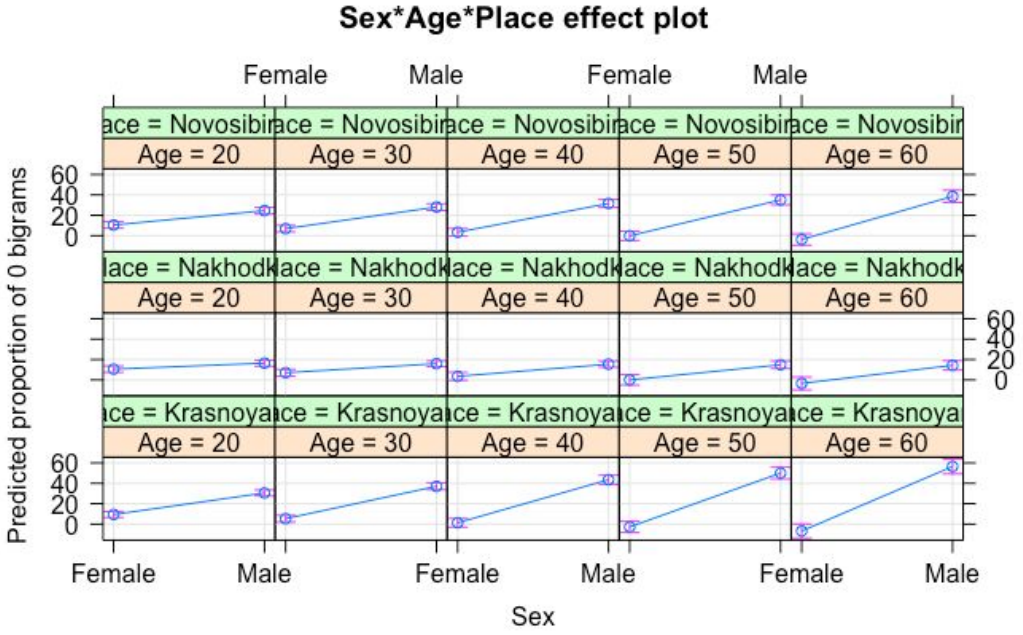


Fig. 10. Effect of age and sex by region on the proportion of “0” bigrams



The last two models allow to hypothesize that there is a major difference between the Siberian cities and Nakhodka: while in Novosibirsk and Krasnoyarsk the differences between speakers of

different sexes are relatively clear, in Nakhodka men and women appear to use pitch more similarly. Another possible interpretation of this result may be that the biological sex in Nakhodka does play a role but our current annotation system does not track these differences.

6. Results

The results of the regression analysis can be interpreted as follows. First, the first unigram-based regression model has shown that male and female speakers use the available pitch range differently. While males mostly use the "Low" part of the range, women use other parts of the range more often. Linguistically, this means that men use less pitch movement in their speech, which may be related to a more expressive function of pitch in male speech (significant pitch changes are rare and therefore more noticeable). The first bigram-based model has shown that male speakers cross the sub-range boundaries significantly less often than females, which supports the hypothesis of the comparably lower pitch use in their speech.

Another possible explanations of these results is that men divide their pitch range differently than women and my version of the tripartite division is not sensitive enough to track the pitch changes. The pitch movement of males may occur within a single sub-range (e.g. within "-1") and the remaining part of the range will be reserved for the rare utterances with an extreme degree of expression. There are thus two possible scenarios: (a) males use pitch movement more rarely than females and (b) the pitch movement in males has a lower amplitude in males but it is not necessarily less frequent. Both interpretations, however, suggest that there is a major difference between male and female speech and only differ in the nature of these differences, which means that the use of the pitch range that can be tracked automatically using a relatively simple technique.

Interestingly, the models with place as a random effect suggest that there is an age-related difference in male speakers with older speakers having more pitch movement. Thus, though the age-related differences are not seen globally, they exist within each areal. From the linguistic point of view, this means that older men use pitch more intensively than younger ones but there is no such difference in women. It may also mean that the parts of the pitch range get re-organised with the increase age and the available range starts to be used differently.

The regional difference between Nakhodka and other cities tracked by the last two models allows to hypothesize the existence of an areal comparative concept, namely of the difference between men and women in the intensity of the pitch use. This means that different regions of

Russia may differ with respect to whether men intonate “less” or somehow else differently than women or not. Another possible interpretation of this result may be that the biological sex in Nakhodka does play a role but our current annotation system does not track these differences. Both results, however, suggest that different regions of Russia do differ with respect to how men and women use pitch.

7. Further Research

The further research that I intend to conduct implies adding new data to the sample (in particular, the data recorded in Izhevsk, Novosibirsk, Yakutsk, and Moldova) and model training using larger datasets, especially using the datasets containing the relative time coordinates. Another promising direction of research is a more detailed study of the intonation in male speakers.

8. References

Chafe, W. L. (1980). The pear stories: Cognitive, cultural, and linguistic aspects of narrative production.

Daniel, M., Dobrushina, N., & Knyazev, S. (2011). Highlander’s Russian: Case Study in Bilingualism and Language Interference in Central Daghestan. *Instrumentarium of Linguistics: Sociolinguistic Approach to Non-Standard Russian. Slavica Helsingiensia*, *40*, 65-93.

Féry, C. (2017). Intonation and Prosodic Structure. CUP.

Kachkovskaia, T., Kocharov, D., Skrelin, P. A., & Volskaya, N. B. (2016). CoRuSS-a New Prosodically Annotated Corpus of Russian Spontaneous Speech. In LREC.

Keijsper, C. E. (2003). Notes on intonation and voice in modern Russian. Studies in Slavic and General Linguistics, 30, 141-214.

Odé, C. (1989). Russian intonation: a perceptual description (Vol. 13). Rodopi.

Šimko, J., Suni, A., Hiovain, K., & Vainio, M. (2017). Comparing languages using hierarchical prosodic analysis. In Proceedings of Interspeech-2017.

Usacheva, M. (Manuscript). Dialogue-focused experiments in the field: advantages and disadvantages (a Permic experience).

Аванесов, Р. И. (1984). *Русское литературное произношение*. Просвещение.

Альмухамедова, З. М., & Кульшарипова, Р. Э. (1980). Редукция гласных и просодия слова в окающих русских говорах:(экспериментально-фонетическое исследование).

Бондарко, Л. В., & Вербицкая, Л. А. (1987). *Интерференция звуковых систем*. Изд-во Ленинградского университета.

Брызгунова, Е. А. (1963). Практическая фонетика и интонация русского языка: пособие для преподавателей, занимающихся с иностранцами. Изд-во Московского университета.

Вольская, Н. Б. (2014). Интонация и языковой контакт: прагматический аспект внутри- и межъязыковой интерференции. XLII Межд. филологическая конференция. Избранные труды. Изд-во СПбГУ.

Высотский, С. С. (1973). О звуковой структуре слова в русских говорах. *Исследования по русской диалектологии. М*, 17-41.

Грамматчикова, Е.В., Князев С.В., Лукьянова Л.В., Пожарицкая С.К., Ритмическая структура слова и место реализации тонального акцента в региональных вариантах современного русского литературного языка. Манускрипт.

Ерофеева, Т. И. (1979). *Локальная окрашенность литературной разговорной речи: учебное пособие по спецкурсу*. Пермский гос. университет им. АМ Горького.

Ерофеева, Е. В. (1997). Экспериментальное исследование фонетики регионального варианта литературного языка. *Пермь: Изд-во Перм. ун-та*, *140*.

Каринский, Н. М. (1929). Язык образованной части населения г. Вятки и народные говоры. Учен. зап. Ин-та яз. и лит, 3, 43.

Князев, С. В. (2006). Структура фонетического слова в русском языке: синхрония и диахрония. *М.: Макс-Пресс*.

Панов, М. В. (1967). *Русская фонетика*. Просвещение.

Пауфошима, Р. Ф. (1983). Фонетика слова и фразы в севернорусских говорах.

Степанова, С. Б., Асиновский, А. С., Богданова, Н. В., Русакова, М. В., & Шерстинова, Т. Ю. (2008). Звуковой корпус русского языка повседневного общения «Один речевой день»: Концепция и состояние формирования. Компьютерная лингвистика и интеллектуальные технологии, (7), 14.

Щерба, Л. В. (2007). Избранные работы по русскому языку. Аспект Пресс.

Янко, Т. (2017). Интонационные стратегии русской речи в сопоставительном аспекте. Litres.

Contact details:

Ilya Chechuro

National Research University Higher School of Economics. School of Linguistics. Assistant/ Linguistic Convergence Laboratory. Junior research fellow. E-mail: ilyachechuro@gmail.com

Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.