Dialectologia 17 *(2016), 1-22.*
ISSN: 2013-2247

Received 15 March 2015.

Accepted 21 July 2015.

DIATECH: TOOL FOR MAKING DIALECTOMETRY EASIER

Gotzon AURREKOETXEA, Gotzon SANTANDER, Iker USOBIAGA & AITOR IGLESIAS

University of the Basque Contry[1]*

gotzonaurre@gmail.com/sunnyworld84@hotmail.com/ikeruso@gmail.com/

txadonak@gmail.com

Abstract

Diatech[1] is a web application to analyze linguistic differences in a quantitative exploration in a friendly way, allowing users even if they have not computational expertise. The application conceived as a tool which provides all tools that a dialectologist needs when his objective is to draw conceptual maps (the occurrence of individual features), to delimitate the linguistic distance between dialects or to draw boundaries in dialect areas. *Diatech* creates different types of maps (isoglotic, beam, similarity maps, etc.), cluster and MDS analyses, it checks out centroid localities and analyzes the linguistic features that provoke the main variation. It uses different linguistic measures as RIV (Relative Index Value) or Levenshtein algorithm, taking into account different types of answers (such as orthographic, phonetic or lemmas). All maps and illustrations can be downloaded in different image format (RGB o CMYK) and sizes. Since the application was launched, the responsible team has continued improving it, making it easier to use and more powerful in the statistic techniques.

Keywords

linguistic variation, dialectometry, computer program, automated maps

* UPV/EHU, Irakaslegoaren Unibertsitate Eskola, Leioa, Bizkaia.

[1] Diatech is accessible at http://eudia.ehu.eus/diatech.

DIATECH: UNA HERRAMIENTA PARA UNA DIALECTOMETRÍA MAS ASEQUIBLE

Resumen

Diatech es una aplicación web para el análisis de las diferencias lingüísticas desde el punto de vista cuantitativo en un entorno amigable, para uso de lingüistas no habituados a usar programas informáticos sofisticados. La aplicación ha sido concebida como una herramienta que proporciona todos los útiles necesarios para el dialectólogo que pretende crear mapas conceptuales (las ocurrencias de características individuales), delimitar distancias lingüísticas entre dialectos o determinar las fronteras de áreas dialectales. *Diatech* puede crear diferentes tipos de mapas (mapas isoglóticos, mapas de rayos, mapas de similitud, etc.), análisis de clúster y de escalamiento multidimensional; puede también determinar las localidades centroides de un área dialectal o detectar las características lingüísticas que provocan la mayor variación. La herramienta usa diferentes medidas lingüísticas, como RIV (Relative Index Value) o la distancia de Levenshtein, teniendo en cuenta diversos tipos de respuestas (ortográficas, fonéticas o lemas). Tanto los mapas como las diversas ilustraciones pueden ser descargados en diferentes formatos (RGB o CMYK) y tamaños. El equipo continúa mejorando e implementando nuevos recursos estadísticos con el objetivo de hacer la herramienta más accesible a los investigadores.

Palabras clave

variación lingüística, dialectometría, programa informático, cartografía automatizada

1. Introduction and background

1.1 Background

The study of dialect boundaries has been carried out since the beginning of dialectology. Fortunately, the discussion about the existence of the boundaries it is over and nowadays it is accepted by all dialectologists. The study about these boundaries has been carried out by the isogloss methods in the traditional dialectology. Nevertheless, this method has been criticized (Kessler 1995, Goossens 1977, Inoue 1996, among others), while other methods have been applied in the last decades.

By quantifying linguistic data and using statistical procedures, modern dialectology has been able to measure linguistic distances between different localities and classify dialects in a more accurate and scientific way. Besides its suitability to measure

differences between dialects, dialectometry shows the analysis of the dialects in a much more attractive way than it use to do in the past.

Jean Séguy (1973) was the creator of the discipline, although he had to deal with quantification of the data in a manual way, by making by hand all quantification task. He was able to measure linguistic distances between close and far localities in a bidirectional way, taking two localities each time and finally drawing the Gascony's features map. In contemporary works such as H. Goebl (1976) and H. Guiter (1973), they started working in a similar way. But it was Goebl who really developed and achieved an enormous progress. And he was the one who conceived the first computing package for quantitative dialectology. This package, named VDM and created by Haimerl (1999), introduced all techniques that Goebl had been developing in the last 30 years. It uses different linguistic distances (RIV and WIV), wide variety of synoptic maps (similarity, transitional and conservative areas, etc.), beam and isoglotic maps, cluster and correlation analysis (Goebl 2010). This desktop program gathers up the tools that a dialectologist needs from the geolinguistic point of view, facilitating users to draw individual features maps and synthetic maps, using quantitative analyses for achieving the mentioned purpose.

The use of computerized statistic packages compels users to learn not only Linguistics but also Statistics, because dialectometry is a multidisciplinary task, in which dialectal data must been analyzed in a quantitative way too.

The second package in dialectometry was developed by Peter Kleiweg's L04 package (www.let.rug.nl/kleiweg/L04). This package was finally developed in the Gabmap online version (Nerbonne *et al.* 2011).

1.2 Motivation

Although we knew the computerized programs for developing dialectometry, we began with the task of creating a new one, because we had two concerns in mind, that they have not been accurately solved yet by those sorts of software: on the one hand, the problem of "multiple responses" and, on the other one, the transportability of the statistical outcomes (Aurrekoetxea *et al.* 2013).

The "problem" of the multiple responses (henceforth MR) is relatively new in dialectology. The traditional dialect atlases do not gather more than one response for each question from one locality (mainly they gather the "best" response, when more than one is collected). Nevertheless, every linguistic knows that language is constantly changing; whether one of the main features of the language is to be changing continuously, in the real situation of any of them, will be cases that show this variation.

Whether the traditional dialect atlases do not gather this intralocality variation is a cause of the used methodology. This subject is changing in modern atlases, in which linguistics use different methodologies, taking into account the need of recording this local variation. This is the case, for example, in the Linguistic Atlas of Basque (henceforth EHHA) (Euskaltzaindia 2010-2016). The treatment of these MRs in the quantitative dialectology has been a constant concern (Aurrekoetxea 2002). The VDM package did not consider this kind of data, and we disagreed with the solution given by Gabmap (Aurrekoetxea *et al.* 2013: 26-27).

The second motivation, the transportability of the statistical outcomes, is more complex and requires improves from theoretical and statistical points of view. In science it is crucial the transportability of the outcomes from one team to another; it is essential for the advance of the science. However, it is not the case of dialectology. Dialectologists have epistemological weakness when some basic concepts have to be defined. Even if there is an amount of definitions about 'dialect', each dialectologist uses it in different ways. It is a clear consequence of the lack of accuracy in the definition of this term. The fact that there are many definitions is an evidence of the lack of accuracy. Now, having improved in the measurement of the linguistic distance or similarities and being accustomed to statistical techniques which allow us to measure these distances in a much more accurate way than with the tools of the traditional dialectology, dialectologists must improve in the data quality and quantity that has to be used.

Taking into account these reasons, we have put our efforts into creating a new package for doing dialectology. Concretely, a software that will continue improving the performance. This version, firstly, provides tools for uploading data to the internet and for drawing the corresponding conceptual maps; secondly, it offers the possibility of managing databases which are hosted into it and of carrying out many dialectometrical

analyses (similarity maps, transition zones, detection of conservative zones, analysing the correlation between two fields, beam and honeycomb maps, — deterministic and probabilistic-cluster analysis —, MDS, centroid localities and the analysis of the linguistic features that provoke the main variation), using string (Levenshtein) and nominal or categorical (RIV (Relative Index Value-RIV) and WIV (Weighted Index Value-WIV)) linguistic distances following different DM schools and taking into account different types of answers (such as orthographic, phonetic or lemmas — Goebl's *taxat*).

2. The main features of *Diatech*

Diatech is based on the main dialectometrical methods developed until nowadays and includes the most useful tools that are described in the scientific literature. Apart from the tools to manage the project (importing and exporting of the data, management of the database (questions, answers, locations and informants) and project, the tool provides a wide range of technical procedures for doing DM, allowing users to carry it out in different steps: selecting the type of answer, linguistic field, linguistic distance, algorithms and analysis type.

2.1 Uploading data to Diatech

The first step is one of the most difficult ones to be overcome by users who are not skilful in the use of computers. In fact, it is very easy, but linguists must know that it is enough to insert one extra comma by mistake in the database or to leave one out and we will not be able to upload the data. Taking into account these difficulties, the team of the tool has decided to help the user by including only one database structure. This database must have three kinds of data: locations, questions and answers.

	location1	location2	location3	...
question1	answer1-locat1	answer1-locat2	answer1-locat3	...
question2	answer2-locat1	answer2-locat2	answer3-locat3	...
question3	answer3-local1	answer3-locat2	answer3-locat3	...
...	...	...	...	...

Figure 1. Structure of the database

The structure of the single file contains these features: the first line should include the locations. In the following lines, the first column should have the translated questions with their corresponding language and the other columns should have the answers as it is shown in Figure 1. Whether there is more than one answer in a locality, the answers will be separated by commas.

In some geolinguistic projects localities can have more than one informant; whether users would like to analyze and compare the data of different informants, they have to create as many databases as informants and upload them separately, but it is different if informants have not sociolinguistic motivations (age, social class, etc.) and has not sense to distinguish the responses of them, answers of different informants must be stored in the same database. That is to say, if the user would need to analyze separately the data of two generations, or data of urban and not urban informants, he must create two databases and upload them to *Diatech* in two separate projects.

The requirements for the data structure are the following ones: it must be coded as UTF8 to avoid problems with special characters; it must be stored as .csv format; and finally, they have to be compressed all together in one ZIP file.

2.2 Creating the map

Once the data has been uploaded, we must create the map, including in it all the localities that we have in the database and drawing the boundary. In order to do this, the *Diatech* tool provides a Google map in which the tool has located the localities. Nevertheless, the name of one locality may sometimes be situated in different parts of the world; in these cases, the user has to deal with it, managing and putting it in the

correct position on the map. Once the marking of the boundary is completed, the machine will automatically create the Voronoi Map (Thiessen polygons) including all the localities of the project. Once this is done, we have completed the storage of our data in *Diatech.*

2.3 Managing the project and data

When the data is uploaded, users can manage not only the data, but also all the features of the project. We have different options for using the tool: to manage the project (invite users to the project, export the database, etc.), to manage the database (changing, correcting, adding or deleting locations, questions and answers), to search for answers (by questions, location, etc.), or to display data in maps. Here, we will consider only three aspects: questions, the linguistic field and the type of answer.

2.3.1 Questions

If we chose "question" we will have the list of all the questions of the database, in the same order as in our previous database system, and beginning with the language in which the questions are written.

2.3.2 Linguistic domain or linguistic field

As far as the "linguistic domain" is concerned, the user has different options: if he has data with more than one linguistic domain or field, he can group them into different projects, such as phonology, noun or verb morphology, syntax, prosody and lexicon; this means that each one can be analysed separately, in different projects, to study each field and finally all the projects can be analyzed together, to measure the total linguistic distance.

2.3.3 Type of answer

Concerning the "type of answer", the *Diatech* tool allows people to use different transcription systems of data: phonetic transcription, orthographic transcription or lemma. If the researcher would like to measure linguistic differentiation with the best possible accuracy he must use the phonetic level of transcription system. But he has the option to use less accurate transcription by using orthographic transcription. Finally, the researcher has also the option to use *lemmas*, to take only the essential features, avoiding accidental differences that could be embedded into the data; for example, lexical data can have features that are not strictly belonging to the lexical field, such as phonetic or pronunciation features, and so on. Each type of answer must be stored and uploaded to *Diatech* in a different project.

2.4 Multiple responses (MR)

Diatech gives a suitable solution to the MR problem in dialectology in order to measure multiple responses, similarity coefficients are calculated, particularly the Dice coefficient. Taken two linguistical identities a and b where the respective set of responses have been A and B the similarity index between the two of them would be defined has:

$$\frac{2|A \cap B|}{|A|+|B|}$$

where $|A|$ is the amount of responses gathered in a, $|B|$ the amount of answers in b and $|A \cap B|$ the amount of answers in common in both locations.

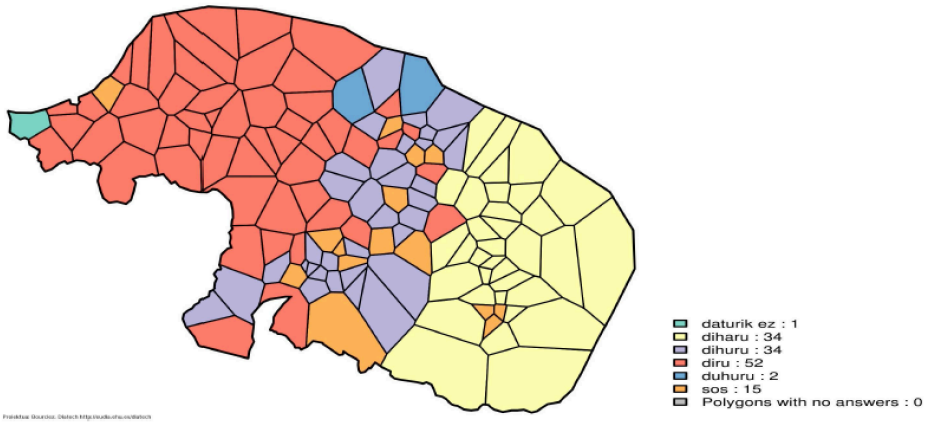
If a dissimilarity distribution is requested the complementary formula applies:

$$1-\frac{2|A \cap B|}{|A|+|B|}$$

Dialectologia 17 *(2016), 1-22.*
ISSN: 2013-2247

2.5 Conceptual maps

Once the database is uploaded and the map is created, the user can draw conceptual maps, as if it was a linguistic atlas; it can be made as many maps as questions have been made. All maps provided in *Diatech* are drawn by using Voronoi polygonation. The conceptual maps, drawn in colours, will be furnished with legend, in which the response will be provided with a determined colour. In Map 1 it is shown the conceptual map referring to 'money' with Bourciez corpus data (Aurrekoetxea, Videgain & Iglesias 2007) (see Map 1). You can see that there are five responses *(diharu, dihuru, duhuru, diru* and *sos*), each one with its colour. The map shows the distribution of each word.



Map 1. 'Money' conceptual map (Bourciez corpus)

3. Doing Dialectometry (DM)

Three kinds of DM can be made by using *Diatech:* unidimensional DM, multivariate aggregate DM and correlative DM. To do all these quantitative analyses, users have to select some basic features as the linguistic distance and the tool provides different algorithms.

3.1 Basic features

Apart from the data type already seen, the tool provides a wide range of types of analysis, linguistic distances and algorithms. *Diatech* supplies the most frequently used similarity measures or linguistic distances in DM; that is nominal or categorical and it is a string measurement. When the analysis is made with lemmas we propose to use categorial measures; among these measurements *Diatech* uses RIV 'Relative identity Value' and WIV 'Weighted Identity Value' (Goebl 1992, 2007, 2010), used by *VDM.* But when the quantitative analysis is made with phonetic or orthographic answers we propose to employ 'Levenshtein distance' (see Nerbonne & Heeringa 2001; Heeringa, Nerbonne & Spruit 2007: 5; Heeringa 2004, for more details), used by *Gabmap*. The researcher has the option to choose one of these, according to the aim of his research. There are more distances: Euclidean, Manhattan distance (Prokic 2009), etc., but there are not widely used in DM and that is why *Diatech* does not provide them.

For the visualisation of the results, *Diatech* gives users the option to use three algorithms: *Med*, *MinMwMax* and *MedMw* (for more information about these algorithms see Goebl 2013).

3.2 Unidimensional DM

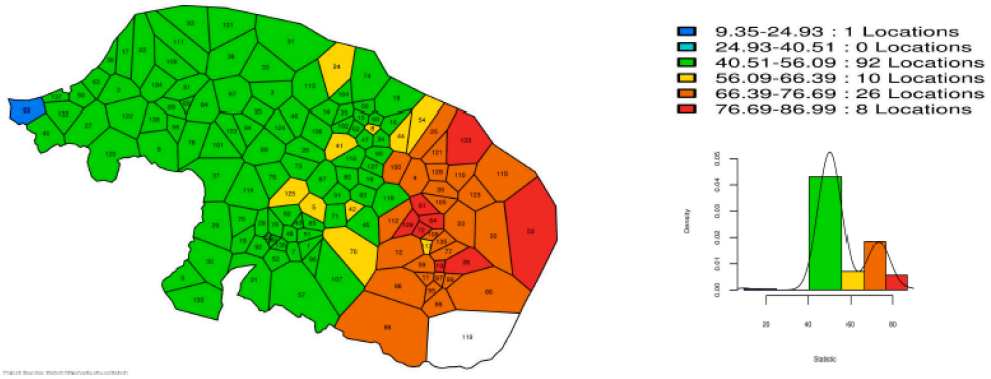
There are three main ways to analyse data by starting with one locality and comparing data from two localities each time: synoptic maps, honeycomb maps and beam maps.

3.2.1 Synoptic maps

The synoptic map in DM is a synthetic map in which the researcher shows the linguistic similarities / dissimilarities among the localities. There are many kinds of synoptic maps: maps of distribution of similarity, maps to show the transitional zones, maps to show conservative areas, maps to show the linguistic center of the dialectal area, etc.

3.2.1.1 Distribution of similarity

Departing always from one locality, the distribution of similarity shows the linguistic difference of each locality respect to the other. The distribution of similarity can change depending on the type of data, the algorithm and the linguistic measure that has been used. Check, for example, the distribution of similarity of Santa Garazi (Map 2), using the *MinMwMax* algorithm.



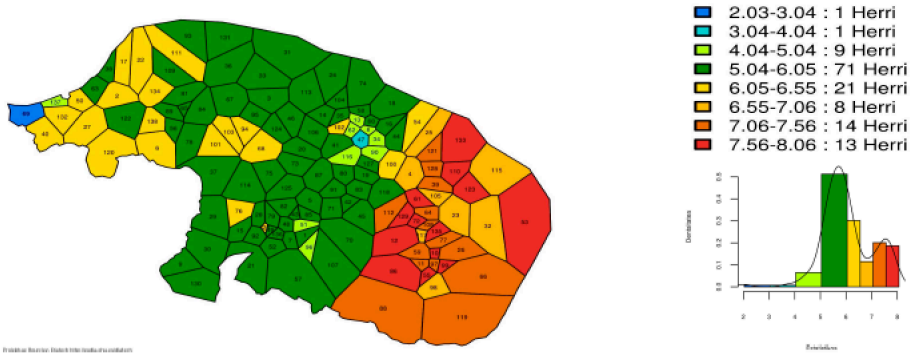
Map 2. Distribution of similarity from Santa Garazi

3.2.1.2 Transitional areas

We consider 'transitional area' an area whose *dialect* has been influenced by the *dialect* of one or more neighbouring local areas. Although many researches have been made about this subject, the concept has been used in different ways: sometimes small areas have been defined as transition areas, but in other cases large areas have been used.

DM, as a technique to measure linguistic distances, has also developed methods for analyzing transitional areas. These areas show a gradual change from one area to another. The standard deviation technique ('synopse des écarts-types' Fr.) is used in DM to detect transitional areas, located between two compact linguistic areas. It is known that the main feature of these areas is that they do not have many of their own features

and that they share the characteristics of their neighbouring localities and areas. The map 3 shows a large transition zone between two linguistic areas, situated one on the West and the other on the East (Map 3).



Map 3. Transitional zone in the Basque Northern area (Bourciez corpus)

3.2.1.3 Conservative areas

To find out where are the more linguistically conservative areas, quantitative dialectology uses the synopsis of skewness ('coefficient d'asymétrie de Fischer' Fr. / 'schiefe' Ger.). This statistical technique is used to analyze the symmetry and its orientation.

Let us assume that we are using a similarity distribution. In general, for the distribution of a certain location (linguistic identity) a positive asymmetry coefficient, a right skewed distribution, would indicate that most of the other locations have low similarity values. On the other hand, if the distribution happens to be negatively skewed, it would indicate that most of the other locations have a high similarity coefficient.

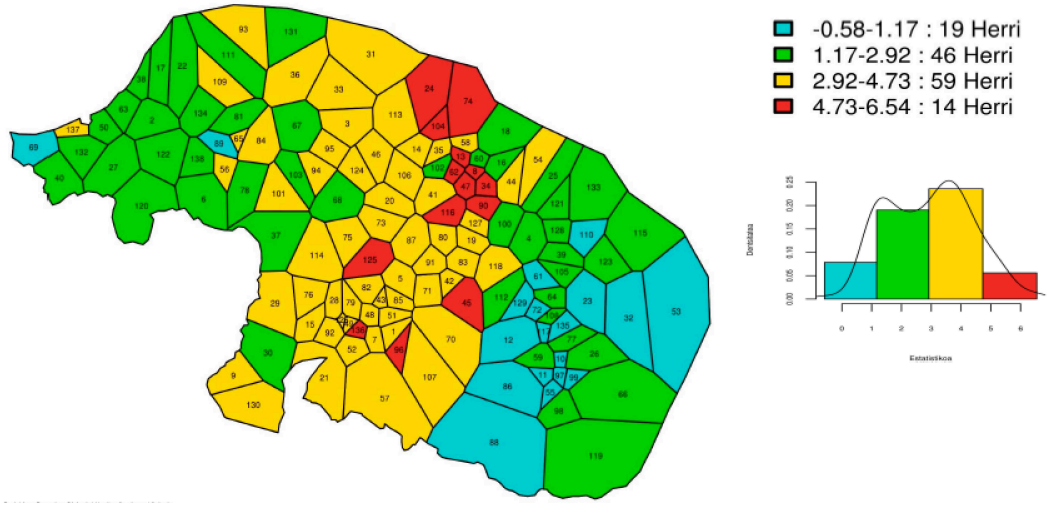
Note that if a dissimilarity distribution is used instead of a similarity one, the meaning of the skewness becomes the opposite.

Taking into account the meaning of this technique two sides can be distinguished: one side has positive values (which fits with linguistically conservative zones) and the other one has negative values (which fits with non-conservative areas).

3.2.1.4

We define 'dialect core' as the group of localities that have the maxima of similarity distributions in the distance matrix of the data. These are the localities which have more similarities with the others. Quantitative dialectology has the 'Synopsis of the maxima of N similarity distributions' called technique to show the largest values of similarity of each locality. Its importance from the point of view of the dialectologist is the detection of the dialect core; that is to say, in these localities have the best relationship with the others in a bidirectional way, taking two localities in every distance. The higher similarity scores that a locality has with its neighbouring localities, the greater the dialectality of the locality is.

In Map 4 there are two groups of localities in red-orange colours: one on the Eastern part (red and orange colours) and the second one on the Western part (just one locality). These localities are extremely close from the linguistic point of view. Between these two groups there is a transitional area.



Map 4. Distribution of maxima of similarities (Bourciez corpus)

3.2.2 Isoglotic map (Honeycomb map)

Using the concept of quantitative isogloss (Goebl 1992), this procedure develops the traditional concept of isogloss and transforms it to include many features (not only one) and doing so in order to enable us to mark more or less important boundaries.

In the isoglotic map every side of the polygons of each locality turns into isogloss. Each side of the polygons takes the value of the linguistic differences between the localities of both sides. Thus, each isogloss has to be drawn with the corresponding characteristic.

3.2.3 Linguistic similarities (Beam map)

While the honeycomb map shows the linguistic differences between localities, the beam map shows the similarities. By drawing maps using this procedure the visualization of the linguistic proximities between localities it is shown and users can easily see the most linguistically homogeneous areas.

3.3 Multivariate DM

One of the most interesting achievements of the DM is the multivariate aggregative analysis. Among the large numbers of technical ways to carry out multivariate analysis in quantitative dialectology, cluster analysis and Multidimensional Scaling (MDS) techniques have been the most widely used.

3.3.1 Hierarchical classification of dialects (Cluster analysis)

There are many classes of clustering; from the Hierarchical, to flat clustering, hard and soft clustering (Prokic 2009), discrete clustering and composite clustering, Bootstrap clustering and Fuzzy clustering analysis, clustering with “noisy” (Dillon & Godstein 1984), etc.

However, whether we ask about the best clustering analysis probably the answer never will be a concrete and specific one. As Prokic (2009) said: "there is no one best clustering algorithm: every algorithm has its own bias […] The success depends on the data set it is used on […] Small differences in input can lead to substantial differences in output".

Diatech uses two types of cluster: the deterministic (hierarchical cluster) and the probabilistic fuzzy cluster. The first one determines the hierarchical classification of the localities, drawing clear cluster of localities, without taking into account the grade of the integration of each locality in the cluster. Meantime, the fuzzy cluster analyzes the grade of integration of each locality in the cluster; thus it draws better the linguistic boundaries of each cluster, showing whether the boundary is abrupt or on the contrary this boundary is not noticeable and more than a boundary is a transitional zone.

Respecting the algorithms used in the cluster analyses, three out of all the possible algorithms have been selected: Ward, Complete and Average, the three most frequently used ones in dialectometry. *Diatech* offers the possibility to choose the group length of the cluster, by selecting in the "group length" section the length which fits best according to the goal of the research. For example; 2 groups them 2 colours and 4 groups, them 4 colours.

Cluster analysis shows two outcomes: dendrogram and corresponding map. The dendrogram shows the hierarchy of the structure of data: grouping data in 6 groups, the green group is the most isolated by far. On the other hand, there is only one locality in dark blue (because of lack of data). It is not time to speak here neither about the best grouping of the dendrogram, nor the dialectal interpretation of these areas, but about that by selecting the group length we can change the configuration of the dialect areas.

3.3.2 MDS analysis

MDS allows us to visualize and simplify datasets with a large number of variables by reducing the dimensionality of the same ones with minimal distortion of the selected distance measure. This is really useful because it makes easier to represent the data in graphical terms.

The chosen particular method for MDS is known as Classical Multidimensional Scaling (MDS) or Principal Coordinates Analysis (PCA). This method is widely used due to the fact, that the loss function minimizes the selected linguistic distance (IRD, IPD, etc.).

The main advantage that PCA has is that observations can still be compared with simple graphical methods without the need of grouping (discretizing) the data. From a distance matrix (in this case 138x138), we are able to simplify another matrix with the selected dimensionality (K = 2, 138x2 in this case). A matrix that is 138 x 138 it is difficult to be expressed graphically; but one that is 138x2 it can be easily represented in a scatter plot.

It is also possible to represent it in a map by combining two basic colors such as green and blue and by adding a higher proportion of green for high values of the first dimension and similar for blue and the second dimension. This makes locations comparable with the need of grouping; therefore, the analyzed region looks 'continuous' rather than grouped. It represents the dialect continua, whereas cluster represents dialect areas.

Analyzing the map from left to right, for this result we can clearly see how there is some kind of linguistic continuum (smooth color changes) across the 'green' part of the map but how, suddenly, a jump occurs and a different area is drawn in blue. The same pattern can be seen in the scatter plot.

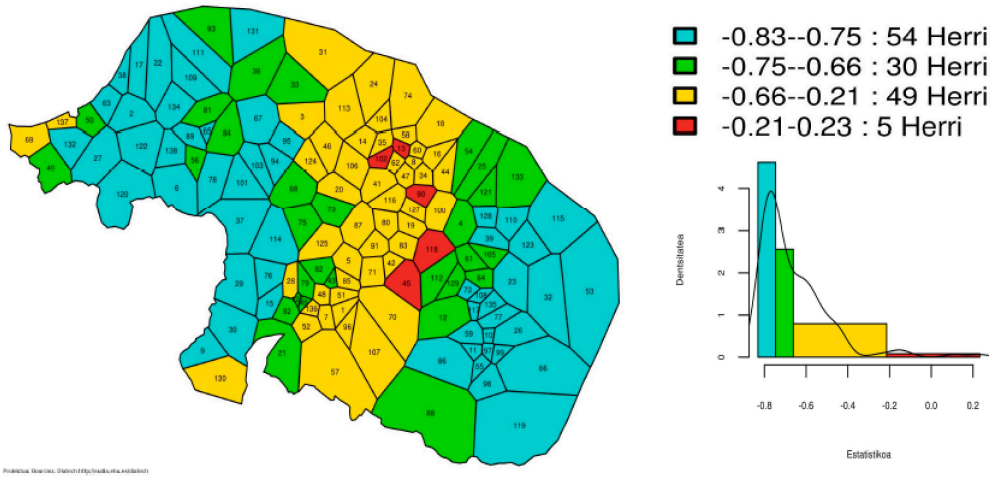
The Mardia measures the loss of the information when the matrix is changed from 138x138matrix to 138x2. The higher loss of information that happens, the lower Mardia coefficient that we will have; the "1" number indicates no loss of information, and "0" way too much loss.

3.4 Correlative dialectometry

One of the most controversial subjects from the very first studies in dialectology has been the relation between the Euclidean distance and the linguistic differences. In effect, Séguy (1971) showed that the biggest Euclidean distance, the biggest linguistic difference. Nevertheless, this equation does not match perfectly in all situations and it

can change from one place to another, because of the humans' management of the space.

Goebl (2005) was the first dialectologist who implemented this kind of analysis into the VDM program, making it easy enough for his followers. *Diatech* has also implemented this technique and allows researchers to analyse the correlation between different dissimilarities. In this case (Map 5), left and right linguistic areas have better correlation with Euclidean distance than the central area has.



Map 5. Correlation between linguistic and Euclidean distance (Bourciez corpus)

"The VDM program also provides r(BP) values. If we look at these values next to the level panel, we see that all of them are positive. The coefficient r(BP) values generally ranges between -1 and +1. According to Aurrekoetxea (2010), whereas the first value indicates a negative correlation between compared variables (in this case lexical similarity and geographical distance), the second one signals a positive correlation. The "0" value indicates absence of any correlation.

3.5. Map, legend and histogram

All maps have a legend and a histogram; they have different information, depending on the map type (analytic or statistic one).

3.5.1 Legend

Legends are automatically created when a map such as an answer map, an analytic map or a statistic map is created. In the first case, legends are filled out with answers (orthographic, phonetic or lemmas) and in the second with intervals produced by visualization algorithms.

3.5.2 Histogram

There are different kinds of histograms. The histograms of *Diatech* show the relative density of each variable. The 'histogram' has three points to be explained: the length (or the height), the width and the line (or the curve). While the measurement of the similarity / dissimilarity (RIV, WIV, etc.) appears in the x-axis (axis of abscissas), in the y-axis (axis of ordinates) the relative frequency is shown, the density. That is to say, the high length (the height), for example, denotes that there are many localities in this section (Figure 2).

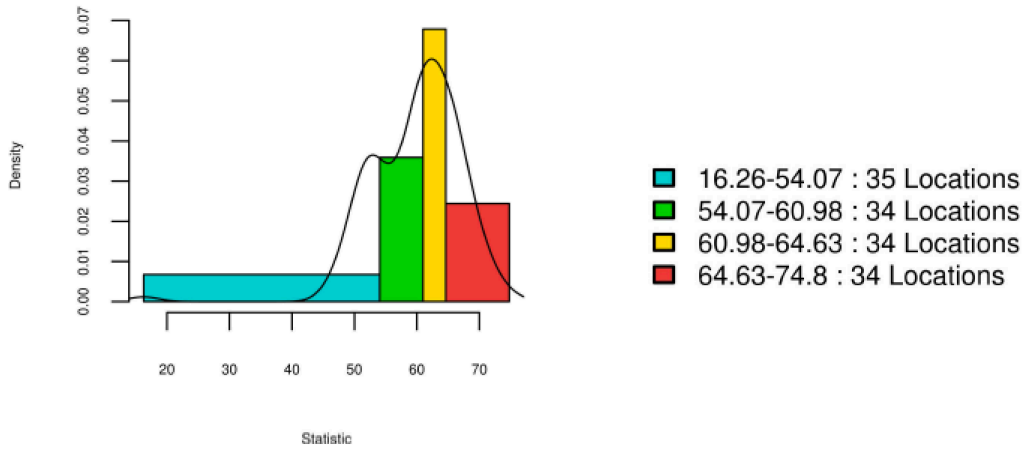


Figure 2. Histogram (Bourciez corpus)

In this case there are 35 localities in the blue area, but as the length between 16,26 and 54,07 is high (37,81), it is broad and it shows low frequency of localities in this group. The width of the green colour is 6,91 and there are 34 localities, so it is thinner but higher (the frequency of the localities is higher); the width of the yellow is 3,65 and

it also has 34 localities (the frequency of the localities is the highest); and, finally, the width of the red colour is 10,17 and there are 34 localities (the frequency of the localities is not very high).

And finally the line on the histogram, a Gaussian kernel estimation of the underlying distribution, indicates an estimation of the best choice of the numbers of groups: the best choice is when the curve and the height of each group agree or are very similar one to the other, by using the minimum of groups. It is a good indicator in order to choose the best grouping of the data.

4. Downloading data and outputs

On the one hand, *Diatech* provides the option to download the distance matrix (numerical distance between localities); and on the other hand, maps, legends and histograms. Underline, that all images have the jpg. format.

4.1 Distance matrix

The distance matrix is calculated in percentages; therefore, users can see the numbers and classify the distances from the smallest to the largest one. It is very important that users of DM come back to the linguistic data and explore which linguistic features have provoked the main distances among localities in different spaces. These distances will be different, according to the linguistic distance chosen, as you can see in the table.

4.2 Images

The *Diatech* tool provides two formats of images: images to be displayed on screens (RGB format) and the ones for editorials (CMYK format). On the other hand, users can also select the size of the images: medium or small. As it has just been mentioned, all images have the jpg. format.

5. Conclusions and future work

With the aim that DM should be available for everybody and socialize among them, *Diatech* puts at the disposal of dialectologists new technical improvements to make dialectometry easier and more comfortable, keeping in mind that well tracked techniques have to be used.

And about future works, the *Diatech* team reasserts that they will work hard to continue improving the tool in the future. We do not get out of our heads one of the linguistic motivations of the creation of *Diatech* tool: the transportability of the statistical outcomes, those we use in dialectology.

References

AURREKOETXEA, G. (2002) "Algunas consideraciones sobre la contrapregunta en las encuestas lingüísticas", in L. Rabassa (ed.), *Mélanges offerts à Jean-Louis Fossa,* Université de Toulouse II-Le Mirail, CerCLid 11/2, 57-65. <http://artxiker.ccsd.cnrs.fr/artxibo-00071400/en/>.

AURREKOETXEA, G. (2010) "The correlation between morphological, syntactic and phonological variation in the Basque language", in B. Heselwood & C. Upton (eds.), *Proceeding of Methods XIII,* Peter Lang, 207-118. <to see coloured maps: https://sites.google.com/site/eudiaehu/home/argitalpenak/argitalpenak-bourciez>

AURREKOETXEA, G., K. FERNÁNDEZ-AGUIRRE, J. RUBIO, B. RUIZ & J. SÁNCHEZ (2013) "DiaTech: A new tool for dialectology", *Literary and Linguistic Computing*, 28(1), 23-30.

DILLON, W. R. & M. GODSTEIN (1984) *Multivariate analysis–Methods and Applications,* New York: Wiley.

EUSKALTZAINDIA (2010-2016) *Euskararen Herri Hizkeren Atlasa,* vol. 1-6, Bilbao: Euskaltzaindia.

GOEBL, H. (1976) "La dialectométrie appliquée à l'ALF (Normandie)", in *Atti del XIV Congresso Internazionale di Linguistica e Dilologia Romanza* (Napoli 1974), Neapel, Amsterdam, vol. II, 165-195.

GOEBL, H. (1992) "Problèmes et méthodes de la dialectométrie actuelle (avec l'application à l'AIS)", in G. Aurrekoetxea & Ch. Videgain (eds.), *Nazioarteko dialektologia biltzarra.*

Agiriak / Proceeding of International Congress on Dialectology, Bilbao: Euskaltzaindia, 429-476.

Goebl, H. (2005) "La dialectométrie corrélative: un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme", *Revue de Linguistique Romane*, 69, 321-355, + 24 maps.

GOEBL, H. (2007) "A bunch of Dialectometric Flowers: a Brief Introduction to Dialectometry", in Ute Smit, Stefan Dollinger, Julia Hüttner, Gunther Kaltenböck & Ursula Lutzky (eds.), *Tracing English through time. Explorations in language variation in honour of Herbert Schendl on the occasion of his 65th birthday,* Austrian Studies in English, v. 95, Wien: Braumüller, 133-171.

GOEBL, H. (2010) "Dialectometry and quantitative mapping", in Alfred Lameli, Roland Kehrein & Stean Rabanus (eds.), *Language and Space. An International Handbook of Linguistic Variation*, *Language Mapping*, vol. 2, Berlin: De Gruyter Mouton, 433-457.

GOEBL, H. (2013) "La dialectometrización del ALPI", in E. Casanova & C. Calvo (eds.), *Actas del XXVI CILFR Congreso Internacional de Lingüística y de Filología Románicas,* Berlin/Boston: Walter de Gruyter, 143-154.

GOOSSENS, J. (1977) *Inleiding tot de Nederlandse Dialectologie. Wolters-Noordho*, Groningen: Wolters-Noordhoff.

HEERINGA, W. (2004) *Measuring Dialect Pronunciation Differences using Levenshtein Distance,* in Groningen Dissertations in Linguistics, 46, <http://www.rug.nl/research/portal/files/9800656/thesis.pdf (last accessed 2014, August 4)>.

HEERINGA, W., J. NERBONNE, J. & M. R. SPRUIT (2007) "Associations among Linguistic Levels" <http://dare.uva.nl/document/98506> (last accessed 2014, August 4).

INOUE, F. (1996) "Computational Dialectology (1)", *Area and Culture Studies*, 52, 67-102.

KESSLER, B. (1995) "Computational dialectology in Irish Gaelic", in *Proceedings of the seventh conference of the European chapter of the Association for Computational Linguistics (EACL)*, San Francisco, CA: Morgan Kaufmann Publishers, 60-66. <arXiv:cmp-lg/9503002> (last accesed 2016, March 3).

NERBONNE, J., R. COLEN, Ch. GOOSKENS, P. KLEIWEG & Th. LEINONEN (2011) "Gabmap: a web application for dialectology", *Dialectologia*, special issue, II, 65-89. <http://www.publicacions.ub.edu/revistes/dialectologiaSP2011/>.

NERBONNE, J. & W. HEERINGA (2001) "Dialect areas and dialect continua" <http://www.let.rug.nl/heeringa/dialectology/papers/lvc01.pdf> (last accessed 2014-August 4).

PROKIC, J. (2009) "Clustering & Bootstrapping" <www.let.rug.n/~nerbonne/teach/rema-stats-meth-seminar/presentations/cluster-boot-prokic-2009.pdf> (last accessed 2014, August 4).

SÉGUY, J. (1971) "La relation entre la distance spatiale et la distance lexicale", *Revue de Linguistique Romane*, 35, 335-357.

Séguy, J. (1973) "La dialectométrie dans l'Atlas Linguistique de la Gascogne", *Revue de Linguistique Romane*, 37, 1-24.

Dialectologia 17 *(2016), 23-49.*
ISSN: 2013-2247

Received 19 September 2014.
Accepted 10 March 2015.

EL USO DE [AW] EN EL ESPAÑOL VASCO: INFLUENCIA DEL ORIGEN FAMILIAR Y LAS REDES SOCIALES

María del Puy CIRIZA
Texas Christian University*
maria.ciriza@tcu.edu

Resumen

Este estudio examina los aspectos sociales que influyen en el uso de características de contacto del español vasco en monolingües, centrándose en: (a) la influencia de las redes sociales bilingües y (b) el origen de los padres distinguiendo entre hablantes cuyos padres son autóctonos del País Vasco y aquellos cuyos padres provienen de otras partes de España. La característica de contacto elegida es la elisión de [ð] con cierre vocálico (cant[aðo] > cant[aw]). El análisis indica que en Bilbao [aw] ocurre más frecuentemente entre monolingües que presentan una alta densidad de hablantes bilingües en sus redes sociales, y aquellos cuyos padres provienen de zonas de alto contacto. Las conclusiones subrayan la necesidad de estudios que investiguen como las redes sociales y familiares influyen en el cambio lingüístico en situaciones de contacto.

Palabras clave

español vasco, [aw], emigración, redes sociales, orígenes familiares

BASQUE SPANISH [AW]: INFLUENCE OF FAMILY ORIGIN AND SOCIAL NETWORKS

Abstract

This study examines the social aspects that influence the use of contact features in Basque Spanish among monolinguals, focusing on: (a) the influence of bilingual social networks and (b) the speakers' parental origin by distinguishing between speakers whose parents are autochthonous from the Basque

* Department of Spanish and Hispanic Studies/2800 S. University Drive, Scharbauer 3100, Fort Worth, TX 76129.

Country, and those whose parents come from other parts of Spain. The contact feature selected is the elision of [ð] with vocalic closure (cant[aðo] > cant[aw]). The analysis shows that in Bilbao [aw] occurs more frequently among monolingual speakers with a high density of bilingual speakers in their social networks, and those whose parents came from high contact areas. The conclusions highlight the necessity of studies that investigate how social and family networks influence linguistic change in language contact situations.

Keywords

Basque Spanish, [aw], emigration, social networks, family origin

1. Introducción

Dentro de los estudios dialectológicos, el concepto de redes sociales ha sido utilizado como alternativa a las investigaciones de lengua y variación en el contexto social (Trudgill 1986; Milroy 1987). A diferencia de los estudios labovianos centrados en separar la variación lingüística en relación con características macrosociales (el estatus social del hablante, su edad o educación), las investigaciones que utilizan la metodología de redes sociales examinan la influencia que el entorno social más cercano al hablante tiene en su actuación lingüística (ver Milroy 1987; Blanco Canales 1989). Desde una perspectiva metodológica, este tipo de estudios recaban información sobre la red de individuos con los que el hablante interactúa habitualmente atendiendo a la densidad (intensidad de la relación entre sus miembros), y multiplicidad de sus redes sociales (número de conexiones entre ellos). A partir de los conceptos de densidad y multiplicidad, el influyente trabajo de Milroy (1987) sostiene que aquellas redes sociales que presentan una estructura densa y múltiple (*strong ties*) se comportan de manera diferente de las que tienen una estructura menos densa y múltiple (*weak ties*): mientras que el primer tipo de redes ejercitan una función de soporte de la norma lingüística, las redes sociales más difusas con lazos de unión más débiles son más propensas al uso de innovaciones lingüísticas.

La teoría de redes sociales ha sido ampliamente utilizada en el campo de la lingüística hispánica tanto en análisis cuantitativos (Blanco Canales 1989; Ávila & Requena 2002) como interpretativos (Cuevas 2001; Villena Ponsoda 2005), y para

múltiples variedades del español peninsular (Blanco Canales 1989, para el español de Alcalá de Henares; Cuevas 2001, Ávila & Requena 2002, Villena Ponsoda 2005, para el español de Málaga; Vann 1996 para el español catalán). En este trabajo nos centramos en el 'español vasco', [1] el dialecto de español con características de sustrato provenientes del contacto español-euskera y analizamos si la estructura de la red social puede actuar reforzando o mitigando el uso de características de contacto. A diferencia de trabajos que siguen la metodología de Milroy (1987) centrada en conocer la red social 'ego' del hablante es decir, investigar cuáles son los lazos más cercanos, el presente estudio analiza las redes sociales 'extensas' con especial atención al contacto de los hablantes con bilingües. La hipótesis principal es que los hablantes *monolingües* pero con extensos lazos sociales con *bilingües* presentarán una mayor frecuencia y número de características de contacto que aquellos informantes con menor densidad de lazos sociales con bilingües (Vann 1996). El respaldo de esta hipótesis lo podemos encontrar en trabajos que han encontrado cómo la constante acomodación lingüística en diferentes espacios sociales (familia, amigos, trabajo etc.) conlleva la adopción de características de contacto (Trudgill 1986; Vann 1996).

El trabajo de campo de este estudio se realizó en dos zonas de contacto: la ciudad de Bilbao con un alto porcentaje de monolingües de español, y Bermeo un pequeño pueblo pesquero con un alto porcentaje de bilingüismo en euskera y castellano. En la zona de bajo contacto de Bilbao se eligieron dos redes sociales: la red social A en la que los hablantes monolingües tenían menos densidad de contacto con bilingües y otra red social B en la que los hablantes todos ellos monolingües también tenían un mayor número de lazos sociales con bilingües. Se eligió también una red social en Bermeo, la red social C, en la que los informantes tenían constante contacto con bilingües.

El estudio se enfoca en una característica propia del español en contacto con el euskera: la frecuente elisión de la aspirante [ð] en el final de palabras con *-ado* (ej. *abogado, cantado, sentado*) que en el español vasco se pronuncia con elevamiento

[1] El presente artículo forma parte de una serie de trabajos en los que se analiza la influencia de las redes sociales en la difusión de características de contacto del español vasco utilizando la metodología de redes sociales. Numerosos son los estudios existentes sobre esta variedad (entre otros, Echaide 1966; Zárate 1976; Urrutia 1988; Etxebarria Aróstegui 2000; González 1999; Elordieta 2003; Isasi, Iribar & Moral del Hoyo 2009; Gómez-Seibane 2012a; Gómez-Seibane 2012b), aunque ninguno de ellos analiza la influencia de la acomodación lingüística con bilingües en las redes sociales.

vocálico [aw].

(1) Cant[aðo] > cant[aw]

Además de estudiar el impacto que la acomodación frecuente con bilingües tiene en el habla de los monolingües en el uso de la variante [aw], analizamos también la influencia del contacto "dialectal intra-familiar" (Potowski 2011). En el País Vasco las migraciones que ocurrieron a mediados del siglo XX atrajeron grandes cantidades de trabajadores de otras zonas de España (Fusi 1984; Ruiz Olabuenaga & Blanco 1994) lo cual pudo acentuar la mezcla dialectal. El objetivo de este trabajo es examinar si existen diferencias en el uso de [aw] entre individuos cuyos padres son autóctonos y aquellos cuyos padres son emigrantes. Es decir, indagamos si, a parte de la red social, es el origen familiar lo que acentúa el uso de las variantes de contacto. La hipótesis es que en zonas de bajo contacto aquellos cuyos padres son de otras provincias de España presentarán una menor frecuencia de [aw] que aquellos cuyos padres son autóctonos ya que los últimos han estado más expuestos a las variantes regionales de contacto.

Numerosos estudios han encontrado como la influencia de la familia es relevante para entender la adquisición de formas locales (Trudgill 1986; Horvath 1998; Kerswill 1996; Kerswill & Williams 2000; Potowski 2011). Kerswill & Williams (2000) encontraron que aquellos informantes que tenían ambos padres nacidos en Londres utilizaban más frecuentemente variantes londinenses que aquellos que tenían solamente un padre de origen londinense y, asimismo, el uso de estas variantes declinaba substancialmente entre aquellos que tenían padres nacidos fuera de Londres. En Norwich, Trudgill (1986) observó cómo los hablantes que no habían nacido en la ciudad y no tenían padres autóctonos adquirían un acento nativo a través del contacto con hablantes locales (con la excepción de formas fonológicas más complejas). En el estudio seminal de Labov (2001) sobre las producciones vocálicas en Filadelfia, el autor halló que el impacto de la red social era más importante que el de la red familiar en el uso de características dialectales.

En el caso del español vasco existen escasos estudios que analicen las consecuencias de la mezcla dialectal así como las diferencias entre emigrantes y autóctonos. En un pequeño estudio con 22 informantes (Ciriza 2009; Ciriza 2015)

descubre que en el caso de la producción de la /r/ con múltiples vibraciones proveniente del contacto con el euskera, el origen familiar tiene una relación clave en su producción: aquellos informantes monolingües cuyos padres son autóctonos procedentes zonas de alto contacto producen la vibrante múltiple con mayor frecuencia que aquellos cuyos padres son de otras partes de España. De manera similar, un estudio anterior Etxebarria Aróstegui (1985) encuentra que el uso de préstamos del euskera en la ciudad de Bilbao es más marcado entre vascos autóctonos monolingües que emigrantes procedentes de otras partes de España.

La pregunta teórica clave es, ¿tiene la familia una influencia relevante en la adquisición de variantes locales y de contacto, o son las redes sociales del individuo las que influencian el cambio lingüístico? Mientras que varias investigaciones demuestran que es la familia el núcleo social de adquisición de variantes locales (Payne 1980; Trudgill 1986; Kerswill & Williams 2000; Potowski 2011), otros estudios confirman que las redes sociales son más relevantes en la adquisición de formas locales (Labov 2001; Vann 1996). El interés de este trabajo es examinar cómo estas dos variables sociales: (1) el contacto con bilingües a través de las redes sociales y (2) el contacto dialectal intra-familiar, influyen en el uso de [aw] en el español hablado en el País Vasco.

2. La variable lingüística

Para estudiar la influencia del contacto dialectal y lingüístico se eligió como característica de contacto la elisión de [ð] en final de palabra con elevamiento vocálico [aw]. Se consideró esta variante porque (a) no únicamente ha sido encontrada en zonas de alto contacto (Echaide 1966), (b) sino también en zonas de bajo contacto, específicamente en la ciudad donde realizamos la mayor nuestro trabajo de campo, Bilbao (Etxebarria Aróstegui 2000). Para el análisis, se consideraron cuatro variantes: (1) la retención de [aðo]; (2) la retención, pero con aspirante relajada [a ^{ð}o]; (3) la elisión [ao]; y (4) la elisión con elevamiento vocálico [aw].

(2) Nombre abogado abog[aðo]> abog[a ^{ð}o] >abog[ao] >abog[aw]

(3) Adjetivos cansado cansa[aðo]>cans[a ^{ð}o]> cansa[ao]>cans[aw]

(4) Participio ha hablado habl[aðo]> abog[a ^{ð}o] >habl[ao]>aboga[aw]

Varios estudios han tratado la elisión de la dental en palabras que terminan en *-ado* tanto en el español peninsular (Zamora Vicente 1967; Lapesa 1981; Samper 1990; Moreno-Fernández 1996) como en el español latinoamericano (López Morales 1980; D'Introno & Sosa 1986; Lipsky 1996). En España, la elisión de la aspirante, [ao], es una característica extendida en la mayor parte de los dialectos del español peninsular (Zamora Vicente 1967; Lapesa 1981), especialmente en Andalucía y Las Palmas (Samper 1990) así como en variedades del centro-norte peninsular (Moreno-Fernández 1996). En España [ao] es utilizada en el registro formal desde comienzos del siglo XX (Díaz Castañón 1975) aunque los manuales de gramática consideran la elisión como un vulgarismo (Seco 1998).

Por otra parte, el uso frecuente de [aw] en el País Vasco se ha relacionado a la frecuencia paralela del euskera de diptongos con secuencias vocálicas abiertas y medias seguidas por vocales cerradas ej. *gauza* 'cosa', *deitu* 'llama' (Etxebarria Aróstegui 2007). El diptongo decreciente [aw] es especialmente marcado en los auxiliares verbales de frecuente uso tales como *dau* 'el está', *eztau* 'no hay'. Interesantemente, esta alternancia ocurre también en los préstamos del castellano al euskera con participios del español que terminan en *-ado.*

(5) Español 'pensado' Euskera bizkaino *pentsau o pentseu*

En el euskera bizkaino, el dialecto hablado en la región donde se desarrolla el presente estudio, Hualde (1991) observa que existe variación dialectal entre las terminaciones [aw] y [ew]; mientras que algunas ciudades y pueblos utilizan [aw], otras (ej. Guernica) utilizan [ew]. Varios análisis cuantitativos indican que [aw] es el diptongo decreciente más frecuente en el euskera de varias localidades bizkainas (especialmente en Bilbao) (Gaminde 2002).

El mayor estudio realizado sobre la distribución social de las variantes de *-ado* ha sido el llevado a cabo por Etxebarria Aróstegui (2000) en la ciudad de Bilbao. El robusto

estudio de la autora se realizó con 72 informantes y se centró en la distribución de las variantes de *-ado* siguiendo líneas labovianas. La autora encontró que la variante no elidida [aðo] era más frecuente en el español hablado en Bilbao produciéndose en más de la mitad de los datos (54%) mientras que las variantes elididas [ao] (19%), y [aw] (17%) eran menos frecuentes. La variante [aw] ocurría más comúnmente en nombres (17%), seguida por adjetivos (12%) y menos frecuentemente en verbos (6%). En cuestión de edad, los hablantes mayores de 60 años utilizaban más [aw] (31%), mientras que los hablantes más jóvenes (19%) y la segunda generación (18%) la favorecían menos. Los resultados de la autora indicaban que los hombres tendían a pronunciar más frecuentemente la variante no elidida [aðo], mientras que las mujeres producen más variantes elididas. Por último, la división sociolectal mostró que tanto los hablantes de la clase social media como los hablantes de la clase social alta presentaron un mayor uso de la variante no elidida [aðo] que de las variantes elididas, utilizando [aw] sólo un 16%. Por otra parte, las clases sociales más bajas tuvieron una mayor frecuencia de variantes elididas especialmente [aw] produciéndose un 42% de las veces.

El estudio de Etxebarria Aróstegui es un punto de importante para analizar el uso de [aw] en zonas de bajo contacto. El presente estudio se centrará en el entorno lingüístico del hablante en sus redes sociales, así como en el contacto intrafamiliar.

3. Contexto histórico

Para entender la distribución geográfica de formas de contacto en el español vasco debemos primero entender los procesos sociales que han afectado el contacto de lenguas en el territorio, dando lugar a la distribución dialectológica actual. Hoy en día, el euskera es hablado en las tres regiones que forman la Comunidad Autónoma Vasca (Alava, Bizkaia, Gipuzkoa) y Navarra, aunque estudios históricos sostienen que en el siglo XVI el euskera se extendía en territorios más al sur en La Rioja, Burgos y partes de Aragón (Mitxelena 1977). La pérdida del euskera se acentúa a mediados del siglo XVI cuando el español gana prominencia como la lengua de los territorios de la Corona, y seguidamente en el siglo XVIII cuando la monarquía Borbónica declara el español como

lengua nacional (Mitxelena 1977). Para la mitad del siglo XVIII, el euskera había desaparecido de numerosos territorios en Navarra y Alaba, aunque seguía siendo ampliamente utilizado en las regiones de Bizkaia y Gipuzkoa (Mitxelena 11977).

Para finales del siglo XIX y principios del siglo XX el uso del euskera ya había disminuido especialmente en las urbes vascas. Paralelamente, las ciudades observan un crecimiento importante debido a las sucesivas olas de emigración de trabajadores procedentes de otras regiones de España. Según los censos, a principios del siglo XX el 26% de la población vizcaína había nacido fuera de la provincia concretamente de provincias aledañas al País Vasco (La Rioja, Burgos y Logroño) y en menor medida de regiones más distantes (Extremadura, Andalucía y Galicia) (Fusi 1984; Ruiz Olabuenaga & Blanco 1994). En Bilbao las primeras migraciones tuvieron como grupo importante los mismos habitantes de la provincia de Bizkaia procedentes de zonas rurales, probablemente bilingües, aunque los trabajadores de otras provincias españolas especialmente Burgos, Cantabria y Valladolid formaban, en conjunto, un grupo mayor (Ruiz Olabuenaga & Blanco 1994). Aunque la emigración sufrió un estancamiento durante la Guerra Civil (Iraola & Zabalo 2013), de 1950 a 1970 se produjo un flujo mayor que el anterior produciendo un espectacular incremento de la población de 1,5 en 1950, a 2,3 millones (Ruiz Olabuenaga & Blanco 1994). La mayoría de los emigrantes se instalaron en las provincias de Bizkaia y Gipuzkoa donde, a la misma vez, el número de hablantes de euskera declinó precipitadamente pasando de ser un tercio a ser un quinto de la población (Clark 1981a: 85; Payne 1975: 104).

Si consideramos los datos demográficos de principios de la democracia observamos la diversidad de origen de los habitantes en el País Vasco. Según la encuesta post-referéndum llevada a cabo por el sociólogo Juan José Linz en 1986 el 37% de los habitantes del País Vasco habían nacido fuera de la región; 8% tenían un padre proveniente de otras provincias de España; 44% tenían ambos padres nacidos en el País Vasco, y el 9% había nacido en el País Vasco pero de ambos padres provenientes de otras provincias de España. De los 2,1 millones de habitantes que habitaban la provincia vasca la mayoría tenían el español como su primera lengua y un 21% hablan el euskera (Linz 1986). Es decir, en cuestiones de autoctonía, el hablante de español era emigrante, hijo de padres emigrantes, o autóctono vasco que perdió o nunca aprendió la lengua

vasca. Por otra parte, el hablante de euskera tenía más de 50 años y vivía en zonas rurales relativamente dispersas de menor población, mientras que el hablante de español residía en zonas urbanas donde se concentraba la mayor parte de la población. En estas zonas urbanas el 40% de los habitantes tenían ancestros vascos, en contraste con el 85% de los habitantes de zonas rurales (Linz 1986).

La situación lingüística del euskera en el País Vasco ha cambiado drásticamente desde la revitalización lingüística y gracias al sistema educativo bilingüe. Hoy en día el número de hablantes de euskera sigue creciendo establemente en la CAE (Comunidad Autónoma de Euskadi) y, según las últimas encuestas sociolingüísticas, el 44,3% de los hablantes son bilingües, 36,4% son bilingües pasivos y un 19,3% de monolingües de español (V Mapa Sociolingüístico 2011). Recientes estudios también apuntan que las nuevas generaciones utilizan más el euskera y tienen actitudes positivas hacia su uso y su revitalización (V Mapa Sociolingüístico 2011).

Los resultados del contacto dialectal en el español entre emigrantes y autóctonos siguen sin embargo sin ser analizados. ¿Es diferente el habla de aquellos cuyos padres son autóctonos en comparación con aquellos provienen de otras zonas de España? Como hemos citado anteriormente, varios estudios muestran la influencia de la familia en la adquisición de variantes (Kerswill & Williams 2000; Trudgill 1986). Considerando que las migraciones pueden tener efectos profundos en los desarrollos lingüísticos, en este estudio compararemos la producción de [aw] en hablantes con diferentes orígenes familiares.

4. Metodología

4.1 Ubicaciones del estudio

Para realizar este estudio se seleccionaron dos ubicaciones en la provincia de Bizkaia, en la que un 30,3% son bilingües, un 20,8% son bilingües pasivos, y un 48,8% son monolingües de español (V Mapa Sociolingüístico 2006).

Bilbao: es la ciudad más habitada del País Vasco y los datos estadísticos muestran

que un 25% son hablantes bilingües, un 22,3% son bilingües pasivos, y un 54,8% son monolingües de español (V Mapa Sociolingüístico 2011). Aunque en Bilbao el número de *euskaldunberris* ha crecido en los últimos años, estudios longitudinales reportan que el 'uso del euskera' es menor que el 'conocimiento académico.' Los estudios más recientes indican que un 3,2% de la población utiliza el euskera en sus interacciones públicas (VI Medición del uso del euskera en la calle 2011).

Bermeo: es una comunidad relativamente pequeña (17.144 habitantes) localizada 30 kilómetros al Este de Bilbao en la costa Atlántica. En la comarca de Busturialdea, donde se incluye Bermeo, el conocimiento de euskera es alto alcanzando el 76% de bilingüismo, 12% de bilingüismo pasivo en euskera y el 13% de monolingüismo español (V Mapa Sociolingüístico 2011). En cuanto al uso del euskera en la calle, los estudios realizados en Bermeo indican que el uso del euskera entre las poblaciones mayores de 65 años es de un 60% y en las poblaciones entre 25-64 años es de un 30% (Bermeoko Udala 2011).

Siguiendo la tipología de lenguas en contacto de Trudgill (1996), mientras que en Bermeo anticiparemos un mayor uso de la variante [aw], en Bilbao encontraremos posiblemente un menor uso de la variante.

4.2 Informantes

En este estudio se recogieron datos de 42 informantes de 3 redes sociales diferentes, dos en Bilbao y una en Bermeo. Tanto los informantes de la red social A (14 informantes) como los de la red social B (14 informantes) eran monolingües. En Bilbao, los sujetos de la red social A tenían menor contacto con bilingües que los individuos de la red social B. Se eligieron específicamente informantes nacidos entre 1970-1990 en el País Vasco para comparar la segunda generación, es decir, los hijos de los inmigrantes cuyos padres llegaron al País Vasco durante los años 50 o antes. En cuestiones de autoctonía, los informantes de la red social A tenían padres de origen emigrante o eran autóctonos de Bilbao; los de la red social B tenían en su mayoría padres autóctonos (véase la Tabla 1). Debemos recalcar que los miembros de la red social B pertenecían a una asociación cultural vasca con una gran identificación con la lengua y cultura vascas.

Para comparar la distribución de [aw] geográficamente se utilizó la red social C de Bermeo (14 informantes) en la que todos los informantes eran bilingües, y contraían a su vez múltiples y densos lazos con otros bilingües.

Por lo tanto, mientras que el nivel geográfico Bermeo y Bilbao constituían dos áreas de alto y bajo contacto respectivamente, en el nivel micro social las redes sociales A, B, C presentaban también un continuo en cuanto al contacto con bilingües. Al final de este continuo encontraríamos la red social C con informantes que tienen una red social bilingüe extensa y que a su vez viven geográficamente en una zona de alto contacto.

4.2 Variables Sociales: El origen de los padres y las redes sociales

En situaciones de contacto Kerswill afirma que "las relaciones sociales entre los diferentes grupos lingüísticos son cruciales para entender los cambios de la lengua" (2006: 1, *mi traducción*). Para estudiar estas relaciones lingüísticas, en este estudio se diseñó una metodología en la que se pudiera medir "la exposición relativa que el hablante tiene con bilingües dentro de su red social" (Vann 1996). Este diseño metodológico se asienta en el trabajo realizado por Vann (1996) sobre el español catalán y a su vez esta metodología ha sido utilizada en un estudio anterior por Ciriza (2009) para la /r/ vasca. El estudio de Vann (1996) examina los aspectos que influyen la producción de innovaciones de contacto particularmente la transferencia pragmática del catalán en el uso de deícticos (los demostrativos 'este', locativos 'aquí', y los usos innovadores de verbos de movimiento 'venir', 'traer'). El autor considera dos factores que pueden contribuir en el uso de estas innovaciones: (1) la asociación de estas variantes al capital simbólico del catalán y/o (2) el hecho de que el hablante tenga extensos lazos sociales con bilingües. Para ello, el autor recogió datos de 52 informantes provenientes de dos redes sociales disimilares: una red pro-catalanista con extensas redes sociales bilingües y otra red con menor número de lazos sociales bilingües. Si bien el autor encuentra que la integración del informante en la red social pro-catalanista no afecta el uso de variantes de contacto, la variable dependiente "exposición relativa al catalán" tienen un efecto significativo en uso de innovaciones procedentes de contacto, es decir, aquellos informantes que tienen lazos extensos con hablantes bilingües

presentan un mayor uso de las variantes de contacto.

Al igual que Vann (1996), en este estudio se realizaron una serie de cuestionarios para identificar al hablante y sus redes sociales. En el primer cuestionario se recogía información sobre (1) el origen de nacimiento de los familiares del informante (padres y abuelos), (2) la lengua principal hablada en casa (con los padres y hermanos) y (3) el modelo lingüístico de escolarización.

Después de recoger los datos se encontró que los padres de los informantes podían ser de: (1) origen otras provincias de España o ser (2) autóctonos del País Vasco. Además, estos dos grupos fueron subdivididos entre aquellos informantes que tenían padres autóctonos de una zona de (a) bajo contacto (la mayoría del Gran Bilbao), y aquellos que provenían de una (b) zona de alto contacto. Teniendo en cuenta estos factores en el análisis se creó un índice que media la "exposición al español vasco a través de los padres."

0= si los dos padres eran otras provincias de España

1= si uno de los padres nació en Bilbao o en otra zona de bajo contacto

2=si uno o los dos padres nacieron en una zona de alto contacto

3=si el hablante era de Bermeo y sus dos padres nacieron en una zona de alto contacto.

Es decir, mientras que el "0" representaría que ambos padres son otras provincias de Españas, el "1" indicaría que uno de los padres nació en Bilbao o en otra zona de bajo contacto; el "2" marcaría que uno o los dos padres nacieron en una zona de alto contacto y "3" reflejaría que el hablante es de Bermeo y sus dos padres nacieron en una zona de alto contacto.

El estudio de las redes sociales del hablante tenía como objetivo examinar como "el grado de exposición relativa del informante a hablantes bilingües" afectaba el uso de la variante. Utilizando el estudio de Vann (1996: 228) como punto de referencia, se desarrolló un cuestionario de 14 preguntas relacionadas con los espacios más frecuentes de interacción social.

HOGAR

1) ¿Tienes tres o más familiares en tu barrio que hablen en euskera?
2) ¿Hablas con ellos en euskera?
3) ¿Sales con ellos frecuentemente?

AFILIACIONES

4) ¿Perteneces a alguna sociedad que defienda la cultura vasca?
5) ¿Hablas en euskera cuando estás en esas afiliaciones?
6) ¿Sales con gente de esa organización que sea nativo-hablante de euskera?

TRABAJO

7) ¿En general trabajas con más gente que habla español o euskera?
8) ¿En el caso en el que trabajes con gente que hable euskera sales con esa gente también?

AMIGOS

9) ¿Tienes tres o más amigos íntimos que hablen euskera?
10) ¿En qué lengua hablas con ellos?
11) ¿Sales normalmente con el mismo grupo de amigos?
12) ¿Sales normalmente con el mismo amigo que es vasco-parlante?
13) ¿Sales normalmente con tres amigos que hablen en euskera?
14) ¿Sales normalmente con cinco amigos o más amigos que hablen en euskera?

Mientras que las preguntas 1, 2 y 3, medían el contacto del informante con hablantes de euskera dentro de la familia, las preguntas 4, 5 y 6 se centraban en espacios culturales vascos, las 7 y 8 examinaban el contacto con bilingües en el trabajo, y las 9-14 en las amistades. Diez de las catorce preguntas fueron codificadas para medir la densidad de las redes bilingües del hablante, es decir, el número de lazos sociales que el informante contrae con hablantes de euskera/español (1, 2, 5-7, 9 y 10-14), mientras

que el resto (las preguntas 3, 6, 8 y 11) se focalizaron en la multiplicidad de las redes sociales (ej. si el hablante tiene un lazo que comparte, es decir si tiene una amistad bilingüe con la que también trabaja). Además, de las 10 preguntas que median la densidad de las relaciones del hablante tres de ellas se centraron en el uso del euskera del informante dentro de la red social (2, 5, y 10) mientras que el resto tenían que ver con el contacto con hablantes de euskera. Seguidamente, las preguntas sobre densidad fueron codificadas para formar un índice que describía la "exposición del hablante a bilingües de euskera/castellano" (Vann 1996: 228); las preguntas de multiplicidad no fueron codificadas aunque fueron utilizadas durante el análisis cualitativo para determinar más concretamente los lazos sociales del hablante.

La codificación de la red social del hablante fue la siguiente: "1" para cada pregunta que tuvo una respuesta positiva con respecto a "exposición al euskera" o "uso del euskera", y "0" cuando el hablante respondía negativamente. Dependiendo del número de respuestas del hablante se creó un índice ordinal para clasificar a los informantes: con baja densidad (0); densidad media; (1) a densidad alta (2) dependiendo del número de respuestas. Finalmente, se dio un índice 3 para los bilingües de Bermeo ya que contraían un menor número de redes bilingües.

0 = densidad baja (para aquellos hablantes que tuvieron entre 0-3 respuestas positivas)

1 = densidad media (para aquellos hablantes que tuvieron entre 4-6 respuestas positivas)

2 = densidad alta (para aquellos hablantes que tuvieron entre 7-10)

3= densidad alta en Bermeo (ya que contraen un mayor número de lazos sociales bilingües a parte de los investigados en el estudio).

De acuerdo con este índice, "0" indicaría una baja densidad, mientras que 1 mostraría una densidad media de contacto con otros bilingües de español/euskera, y 2 implicaría un mayor número de contacto con hablantes bilingües.[2] Por último, el índice 3

[2] En Bermeo, se les preguntó a los informantes también si realizaban viajes frecuentes por cuestiones de trabajo a ciudades como Bilbao de bajo contacto, o si el hablante se relacionaba más frecuentemente con hablantes monolingües que con bilingües. El objetivo era observar que hablantes de Bermeo contraían un

fue creado para los hablantes de Bermeo ya que no únicamente tienen una densidad más alta de hablantes bilingües en sus redes sociales, sino que también viven en una zona de alto contacto.

4.3 Codificación de los informantes

El cuadro (1) representa la codificación de los 42 informantes teniendo en cuenta tres parámetros sociales: (a) origen de los padres; y (b) densidad de hablantes bilingües en las redes sociales, (c) género. Todos los hablantes comprendían edades de entre 26-35 años.

Número	***Nombre***	***Índice-Red Social***	***Origen de los padres***	***Género***
Red Social A-Bilbao				
INF 1	S.V.	1	Bajo contacto	Hombre
INF 2	N.S.	0	Otras provincias de España	Hombre
INF 3	P.D.	0	Otras provincias de España	Hombre
INF 4	S.K.	1	Otras provincias de España	Mujer
INF 5	S.H	1	Bajo contacto	Mujer
INF 6	Z.N	0	Otras provincias de España	Mujer
INF 7	C.V	1	Bajo contacto	Mujer
INF 8	T.C.	0	Otras provincias de España	Mujer
INF 9	A.T.	0	Otras provincias de España	Mujer
INF 10	J.S.	0	Otras provincias de España	Mujer
INF 11	J.C.	1	Bajo contacto	Hombre
INF 12	L.U.	1	Bajo contacto	Hombre
INF 13	S.T.	1	Bajo contacto	Hombre
INF 14	C.T.	0	Bajo contacto	Hombre
Red Social B-Bilbao				
INF 1	C.P.	2	Alto contacto	Hombre
INF 2	S.A.	2	Alto contacto	Hombre
INF 3	C.L.	2	Alto contacto	Mujer
INF 4	F.S.	1	Otras provincias de España	Mujer
INF 5	L.A.	2	Alto contacto	Hombre
INF 6	S.T.	2	Alto contacto	Hombre

alto número de redes sociales monolingües fuera de su red social principal y si esto afectaban el uso de la variante.

INF 7	C.G.	1	Otras provincias de España	Mujer
INF 8	T.A.	2	Alto contacto	Hombre
INF 9	B.M.	2	Bilbao	Hombre
INF 10	S.Z.	1	Bilbao	Mujer
INF 11	L.V	2	Alto contacto	Mujer
INF 12	Z.E	2	Alto contacto	Mujer
INF 13	A.H.	2	Alto contacto	Mujer
INF 14	C.G.	2	Bilbao	Hombre
Red Social C-Bermeo				
INF 1	J.I	3	Alto contacto	Hombre
INF 2	N.R.	3	Alto contacto	Hombre
INF 3	L.S.	3	Alto contacto	Mujer
INF 4	C.U.	3	Alto contacto	Mujer
INF 5	S.T.	3	Alto contacto	Mujer
INF 6	D.F.	3	Alto contacto	Mujer
INF 7	G.M	3	Alto contacto	Mujer
INF 8	Z.C.	3	Alto contacto	Hombre
INF 9	B.S.	3	Alto contacto	Hombre
INF 10	I.L.	3	Alto contacto	Hombre
INF 11	G.L	3	Alto contacto	Mujer
INF 12	I.S.	3	Alto contacto	Hombre
INF 13	T.C.	3	Alto contacto	Hombre
INF 14	N.B	3	Alto contacto	Hombre

Tabla 1. Informante, codificación del nombre, índice de red social, origen de los padres.

Es importante recalcar que todas las amistades de cada una de las redes sociales eran recíprocas y muchas de ellas habían sido consolidadas durante la niñez dando lugar, tanto en la red social A como en la red social B, a relaciones de alta densidad y multiplicidad. Al comparar las dos redes de Bilbao, A y B, primer aspecto que podemos observar es que los informantes de la red social A tenían un menor número de conexiones con hablantes bilingües (7 individuos con índice "0" y 7 individuos tienen índice "1") que los de la red social B (11 individuos con índice "2", y 3 con "1"). En cuanto a origen de los padres, la muestra de la red social A era también equilibrada: 6 informantes con al menos un padre proveniente de una zona de bajo contacto (la mayoría de zonas aledañas al Gran Bilbao o de Bilbao mismo), y 8 con dos padres provenientes de otras provincias de España. En la red social B, la densidad de exposición de los informantes a hablantes bilingües era más alta (9 informantes con una densidad alta, y 3 con una densidad media). Es importante recalcar que todos los informantes de

Bilbao eran hablantes monolingües de castellano en su entorno familiar, y aunque algunos de ellos habían estudiado en modelos bilingües, aunque este aspecto no fue tabulado en los datos. Por último, en la red social C todos los individuos eran bilingües con padres nacidos en una zona de alto contacto y amplias redes sociales bilingües.

5. Resultados

Los datos fueron recogidos a través de entrevistas semiestructuradas centradas en el tema de la identidad vasca, y algunas de ellas relacionadas con el dialecto de español hablado en el País Vasco. Para cada informante, se eligieron 30 producciones de palabras con terminación *-ado* (un total de 1260 producciones), y seguidamente estas fueron separadas dependiendo si el hablante había pronunciado (1) [aw]; (2) [ao]; (3) [a ^{ð}o];o [aðo]. La Tabla (2) representa los resultados por red social.

	[d]	[ð]	[ao]	[aw]
Red Social A	31%	52%	11%	6%
	(129)	(218)	(47)	(26)
Red Social B	39%	53%	5%	2%
	(165)	(224)	(23)	(8)
Red Social C	59%	31%	8%	2%
	(248)	(132)	(32)	(8)

Tabla 2. Resultados por red social.

Como podemos observar las tres redes sociales muestran una tendencia hacia el uso de variantes elididas, aunque si bien la tendencia en Bermeo es hacia la variante elidida [aw] (59%), en Bilbao la tendencia es hacia [ao] (52%, 53%). Estos datos son diferentes a los de Etxebarria Aróstegui (2000) recogidos en Bilbao la cual encuentra una tendencia mayor hacia variantes no elididas especialmente [aðo] produciéndose en más de la mitad de sus datos (54%), mientras que las variantes elididas [ao] (19%) y [aw] (17%) son menos frecuentes. Por el contrario, en nuestros datos de Bilbao la tendencia

es hacia las variantes elididas (un 83% de los casos en la red social A, y en un 92% de los casos en la red social B), mientras que variantes no elididas se producen en menor medida (17% en la red social A; 7% en la red social B). Al final de este continuo encontraríamos la red social C con una mayor tendencia hacia variantes elididas (90%), e incluso menor que en Bilbao hacia las no elididas (10%).

Orientaremos ahora nuestro análisis hacia una comparación de la producción de [aw] en Bilbao con respecto al índice que creamos para medir "la densidad de hablantes bilingües en las redes sociales del hablante." Para realizar esta medición, se dividieron los datos de los informantes entre aquellos que tenían una densidad baja (índice "0") una densidad media; (índice "1") y una densidad alta (índice "2"). Los datos fueron analizados a través del programa estadístico SPSS realizando un análisis de varianza ANOVA de factor único con un nivel de confianza de $p < 0,05$. Se siguieron tres métodos de análisis: descriptivo, ANOVA y el test post-hoc de Steel-Dwass-Critchlow.[3]

Grupo	*Número*	*Media*	*Desviación Estándar*
Índice 1	7	6,9	7,37
Índice 2	10	8,85	10,08
Índice 3	10	12,45	13,35

Tabla 3. Estadística descriptiva del uso de [aw] en referencia al índice de *"densidad de hablantes bilingües"*.

Source of Variation	*SS*	*Df*	*MS*	*F*	*Valor-P*
Entre grupos	151,2914	2	75,64569	5,5592	**0,010378126***
Dentro de los grupos	326,5753	24	13,6073		
Total	477,8667	26			

$p < 0,05$

Tabla 4. ANOVA con factor único en referencia al índice de "densidad de hablantes bilingües" y uso de [aw].

[3] Para la elección de los método estadísticos de se siguió el libro de Dörnyey (2007) "Research methods in applied linguistics" el cual arguye que los datos deben ser presentados primero de una manera descriptiva, seguidos por análisis de varianza, y un test post-hoc.

	Índice 0	Índice 1	Índice 2
Índice 0	**1**	0,283	**0,017***
Índice 1	0,283	**1**	0,120
Índice 2	**0,017***	0,120	**1**

p < 0,05

Tabla 5. Test-Post-hoc: Comparación múltiple por pares utilizando el procedimiento Steel-Dwass-Critchlow-Fligner-T apareado/Valores p.

El análisis de varianza demostró que había diferencia entre los informantes con un índice#1 (M = 6,9, DS = 7,37), con un índice "2" (M = 8,85, DS = 10,08) y con un índice "3" (M = 12,45, DS = 13,35), F (2,24) = 26, p < 0,05. El tamaño del efecto es también grande (eta squared = 0,31) indicando que tanto la magnitud del efecto como la diferencia entre grupos es significativa.

Sin embargo, al realizar post-hoc test a través de una comparación múltiple por pares utilizando el procedimiento Steel-Dwass, encontramos que al contraponer grupos no todos los resultados son relevantes. Por ejemplo, no son significativos (*p* = 0,283) los resultados de informantes con un índice de densidad baja ("0") en relación con los informantes de un índice de densidad media ("1"), y tampoco los de una densidad media ("1") con una densidad alta ("2") (*p* = 0,120). Sin embargo, los resultados entre los informantes con un índice de densidad baja con los de una densidad alta sí son significativos (*p = 0,017*). Estos resultados indican que los informantes con una densidad baja producen [aw] de una manera significativamente menor que los de informantes con densidad alta aunque las comparaciones entre los otros grupos no es relevante.

El siguiente análisis consideró las diferencias relacionadas con el origen de los padres distinguiendo 3 índices: índice "0" para individuos con dos padres otras provincias de España; índice "1" al menos un padre/madre proveniente de una zona de bajo contacto; índice "2"= al menos un padre/madre proveniente de una zona de alto contacto. Las tablas 6, 7, 8 presentan los datos descriptivos junto con los análisis de varianza y el test post-hoc de Steel-Dwass.

Origen de los padres	Número	Media	Desviación Estándar
Otras provincias de España	9	7,25	1,04
Zona de bajo contacto	11	10,29	1,24
Zona de alto contacto	9	13,95	1,33

Tabla 6. Estadística descriptiva con referencia al índice "origen de los padres" y uso de [aw].

Fuente de variación	*SS*	*Df*	*MS*	*F*	*Valor-p*	*F crit*
Entre los grupos	178.8625	2	89.43127	6.843531	**0.004656***	3.422132
Dentro de los grupos	300.564	23	13.068			
Total	479.4265	25				

p < 0,05

Tabla 7. ANOVA con factor único con referencia al índice "origen de los padres" y uso de [aw].

	Alto	Bajo	Otras partes de España
Alto	**1**	0,085	**0,008***
Bajo	0,085	**1**	0,323
Otras partes de España	**0,008***	0,323	**1**

Tabla 8. Test-Post-hoc: Comparación múltiple por pares utilizando el procedimiento Steel-DwasCritchlow-Fligner-T apareado/Valores p.

En el caso de la correlación entre el uso de [aw] y el índice "origen de los padres" puede concluirse también, con un nivel de confianza superior al 95%, que para los tres grupos: informantes con padres de otras partes de España (M = 7,25, D.S = 1,04), padres autóctonos procedentes de zonas de bajo contacto (M = 10,29, D.S = 1,24), y padres procedentes de zonas de alto contacto (M = 13,95, D.S = 1,33), los resultados son significativamente diferentes (p < 0,05). El tamaño del efecto (eta squared = 0,68 > 0,14) es también significativo lo cual indica que tanto la magnitud del efecto como la diferencia entre grupos son significativos. Sin embargo, el post-hoc test de Steel-Dwass muestra que sólo en el caso de los extremos del continuo los resultados son relevantes. Es decir, sólo la comparación de resultados entre los informantes con padres "de otras partes de España" y los "autóctonos procedentes de alto contacto" son relevantes (*p = 0,008*), mientras que la comparación de resultados entre el grupo de informantes con padres provenientes de zonas de "bajo contacto" no son significativos, ni tampoco los resultados de "padres de otras partes de España" con los de "bajo contacto". Por lo

tanto, los resultados indican que solo los informantes con padres provenientes de otras zonas de España producen [aw] de una manera significativamente menor que los informantes con padres autóctonos de una zona de alto contacto.

Finalmente, los diagramas de cajas del Gráfico1 muestran las diferencias entre los grupos de "densidad" y "origen de los padres".

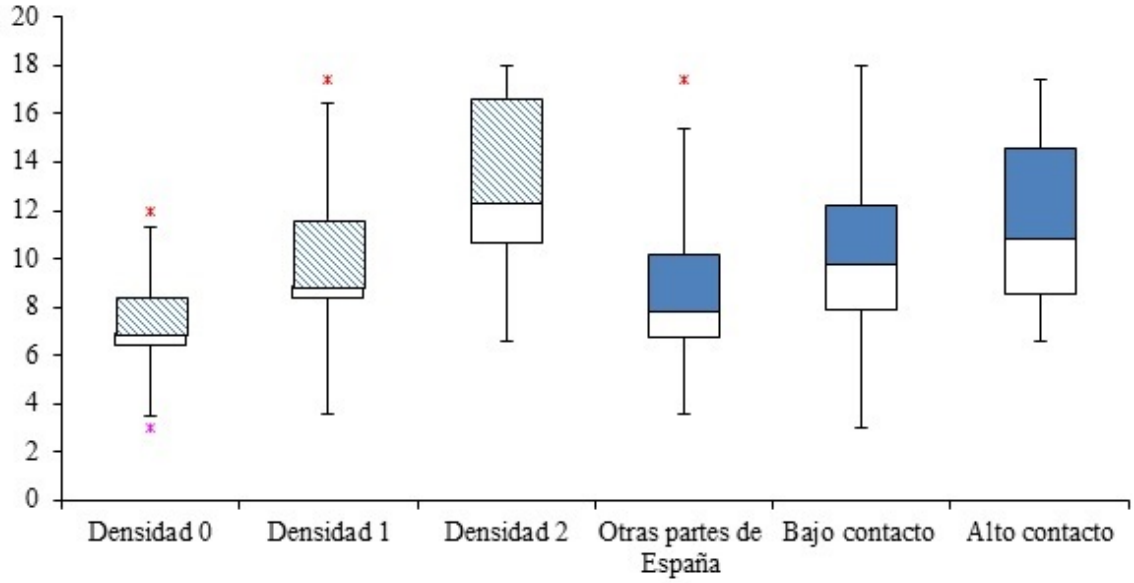


Gráfico 1. Resultados de medias por grupo atendiendo a "densidad" y "origen de los padres".

Podemos ver que las medias de [aw] de los tres grupos que miden la densidad junto con las que miden el origen de los padres son bastante parecidas. Sin embargo, si bien las medias son muy similares, los informantes con una densidad más alta tiende en general a tener resultados mas altos que aquellos con origenes de padres provenientes de zonas de alto contacto.

Finalmente, se realizó un analisis de regresion múltiple para saber si ambos factores juntos la "densidad de hablantes bilingües en la red social x origen de los padres" potenciaban el uso de [aw]. Los resultados de regresion los podemos ver en el cuadro de abajo.

Source	Value	Standard error	t	Valor p	Lower bound (95%)	Upper bound (95%)
Intercept	8,018	0,927	8,652	< 0,0001	6,114	9,923
Origen padres*densidad	1,466	0,389	3,769	**0,001***	0,666	2,265

Tabla 9. Regresión múltiple de densidad x origen de los padres

Si observamos la interacción de las dos variables "origen de padres" por "densidad" podemos observar que el valor es p = 0,001 es significativo. Es decir, el hecho de que un individuo tenga padres de origen autóctono y una alta densidad de redes bilingües, potencia el uso de la variante [aw]. Por el contrario, el hecho de que un individuo sea de origen inmigrante y tenga una red social con pocos lazos bilingües tiene una correlación directa con un menor uso de la variante [aw].

6. Discusión y conclusión

En este estudio se examinó el uso de [aw] en el español vasco con respecto a dos variables sociales: el origen de los padres y la densidad de hablantes bilingües. Para realizar el análisis se siguió un diseño metodológico en el cual se utilizaron cuestionarios para recoger datos sobre ambas variables dependientes. La creación de índices permitió agrupar y tabular los datos estadísticamente de informantes que tenían (1) alta, (2) media y (3) baja densidad de contacto y se agruparon también los informantes cuyos padres provenían de (1) padres autóctonos procedentes de zonas de alto contacto, (2) aquellos que eran autóctonos, pero provenían de zonas de bajo contacto, y (3) los que tenían padres provenientes de otras partes de España.

Los resultados muestran que tanto los informantes de Bermeo como los de Bilbao tienden hacia formas elididas [ao] o [aw] sin embargo, mientras que los informantes de Bilbao tienden hacia [ao] los de Bermeo tienden hacia [aw]. Estos resultados son diferentes a los encontrados anteriormente por Etxebarria Aróstegui la cual descubre un mayor uso de formas no elididas. Estos indican un cambio lingüístico en progreso hacia formas no elididas, y reafirman la hipótesis de que [aw] es una variante de contacto ya

que su distribución es más frecuentemente en la zona de alto contacto en Bermeo. En cuanto al análisis de [aw] en las dos redes sociales de Bilbao se encontró que solamente los resultados que comparaban los informantes con un índice de densidad baja con los de una densidad alta eran significativos. Lo mismo ocurrió con los índices que medían el origen de los padres, donde solo los resultados de los informantes con padres procedentes de otras zonas de España con aquellos que tenían padres procedentes de una zona de alto contacto eran significativos. Es decir, en general las diferencias sólo pueden verse entre los extremos del continuo. Finalmente, el análisis de regresión múltiple muestra que la interacción de las dos variables "densidad x origen de los padres" tienen un impacto significativo en los datos.

Por lo tanto, ¿podemos concluir que es el origen de los padres o la densidad de hablantes bilingües en las redes sociales lo que acentúa el cambio hacia [aw]? Los resultados aquí presentados indican que es la alta densidad de lazos sociales con bilingües lo que acentúa un mayor uso de la variante de contacto [aw]. Además, la combinación de ambos factores, el hecho de que los padres provengan de una zona de alto contacto junto con una alta densidad de redes sociales refuerza el uso de la característica de contacto. Sin embargo, estos resultados no pueden dar una conclusión a esta respuesta debido a que la muestra es pequeña y por lo tanto la fiabilidad es menor. Un corpus mayor es necesario para corroborar los datos aquí presentados y concretar cuál de los dos índices sociales, la densidad de bilingües en las redes sociales o el origen de los padres, es más significativo en el uso de variantes de contacto del español vasco.

Este estudio subraya la importancia de metodologías capaces de analizar las variables sociales que impulsan el cambio lingüístico en situaciones de contacto. Como señala Kerswill (1994) a diferencia de en situaciones monolingües, en situaciones de contacto debemos centrarnos en las interacciones rutinarias del hablante con los diferentes grupos lingüísticos que le rodean. Asimismo, la familia, como red social relevante en la adquisición de formas locales debe ser otro elemento que tiende a ser olvidado. En el futuro, aquellos análisis que refuercen la importancia de ambos factores pueden ser de gran utilidad en los estudios de variación lingüística en situaciones de contacto.

Referencias

ÁVILA, Antonio & Félix REQUENA (2002) "Redes sociales y sociolingüística", *Estudios de Sociolingüística*, 3, 71-90.

BERMEOKO UDALA (2011) *Euskara Bermeon*. Datu sociolinguistikoak. [http://www.bermeo-euskaraz.com/euskara-bermeon/datu-soziolinguistikoak]

BLANCO CANALES, Ana (1989) "Redes sociales y variación lingüística", *REIS: Revista española de investigaciones sociológicas*, 91, 115-135.

CIRIZA-LOPE, María del Puy (2009) "Dialect divergence and identity in Basque Spanish", Tesis de Doctorado en Lingüística, Urbana-Champaign: University of Illinois at Urbana Champaign.

CIRIZA-LOPE, María del Puy (2015) "Family transmission and social networks: insights on a Basque Spanish case", en Kim Potowski & Talia Bugel (eds.), *Sociolinguistic Change across the Spanish-Speaking World: case studies in honor of Dr. Anna Maria Escobar*, Suiza: Peter Lang, 5-31.

CLARK, Robert (1981a) "Basque nationalism in Spain since the civil war", en Charles R. Foster (ed.), *Nations without a State: ethnic minorities in Western Europe*, New York: Praeger Publishers, 75-100.

CUEVAS, Inmaculada (2001) *Variación social, reticular e individual de las consonantes obstruyentes, palatales y dentales en Nueva Málaga*, Tesis de Doctorado en Lingüística, Málaga: Universidad de Málaga.

DÍAZ-CASTAÑÓN, Carmen (1975) "Sobre la terminación *-ado* en el español de hoy", *Revista española de lingüística,* 5, 1, 111-120.

D'INTRONO, Francesco & Juan Manuel SOSA (1968) "Elisión de /d/ en el español de Caracas: aspectos sociolingüísticos e implicaciones teóricas", en Henrietta Cedergren, Rafael Núñez, Iraset Páez Urdaneta & Jorge María Guitart (eds.), *Estudios sobre la fonología del Caribe hispánico,* Caracas: La Casa Bello, 135-162.

DÖRNYEY, Zoltan (2007) *Research methods in applied linguistics*, Oxford: Oxford U.P.

ECHAIDE, Ana María (1966) "Castellano y vasco en contacto: tendencias fonéticas vascos en el castellano de los vascohablantes bilingües", *Boletín de la Real Academia Española*, 46, 513-523.

ELORDIETA, Gorka (2003) "The Spanish intonation of speakers of a Basque pitch-accent dialect", *Catalan Journal of Linguistics*, 2, 67-95.

ETXEBARRIA ARÓSTEGUI, Maitena (1985) *Sociolingüística urbana: el habla de Bilbao*, Salamanca: Ediciones Universidad de Salamanca.

ETXEBARRIA ARÓSTEGUI, Maitena (2000) *Variación sociolingüística en una comunidad bilingüe en Bilbao*, Bilbao: Servicio de Publicaciones de la Universidad del País Vasco

ETXEBARRIA ARÓSTEGUI, Maitena (2007) "Sociolinguistics of Spanish in the Basque Country and Navarre", *International Journal of the Sociology of Language,* 184, 37-58.

EUSKO JAURLARITZA (2011) *V Mapa sociolingüístico*, Departamento de educación, política lingüística y cultura

<http://www.euskara.euskadi.net/r59-738/es/contenidos/informacion/argitalpenak/es_6092/ikuspegi_sozio_linguis.html>.

FUSI, Juan Pablo (1984) *El País Vasco. Pluralismo y nacionalidad*, Madrid: Alianza.

GAMINDE, Iñaki (2002) "Gaurko Bilboko Euskararen Ezaugarriak", en Ana Elejabeitia Ortuondo, Juan Otaegi, Adolfo Arejita, Carmen Isasi Martínez & Nagore Etxebarria (eds.), *El espacio lingüístico. Simposio 700 Aniversario*, Bilbao: Universidad de Deusto, 391-414.

GÓMEZ-SEIBANE, Sara (2012a) "Contacto de lenguas y orden de palabras OV/VO en el español del País Vasco", *Lea,* 34, 5-24.

GÓMEZ-SEIBANE, Sara (2012b) "La omisión y duplicación de objetos en el castellano del País Vasco", en Bruno Camus & Sara Gómez Seibane, *El castellano del País Vasco*, Bilbao: Universidad del País Vasco, 193-214.

GONZÁLEZ, María José (1999) *Immediate preverbal placement: A sociolinguistic analysis of Spanish in the Basque Country*, Tesis de Doctorado en Lingüística Hispánica, Urbana-Champaign: University of Illinois.

HAZEN, Kirk (2002) "The family", en Jack Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *Handbook of language variation and change*, Malden: Blackwell, 500-525.

HORVATH, Barbara (1998) "Finding a place in Sidney: migrants and language change", in Peter Trudgill & Jenny Chesire (eds.), *The sociolinguistic reader,* Nueva York: Arnold, 90-102.

HUALDE, Jose Ignacio (1991) *Basque Phonology*, Londres: Taylor &Francis.

IRAOLA, Iker & Julen ZABALO (2013) "Conflicting nationalist tradition and immigration: the Basque case from 1950-1980", *Nations and Nationalism,* 19, 3, 513-531.

ISASI, Carmen, Alexander IRIBAR & María Carmen MORAL DEL HOYO (2009) "Una transferencia vasca: el seseo de hablantes vizcaínos y guipuzcoanos", *Oihenart,* 24, 201-235.

KERSWILL, Peter (1996) "Children, adolescents, and language change", *Language variation and change*, 8, 2, 177-202.

KERSWILL, Peter (2006) "Migration and language", en Klaus Mattheier, Ulrich Ammon & Peter Trudgill (eds.), *Sociolinguistics/Soziolinguistik, An international handbook of the science of language and society*, Berlín: De Gruyter, 1-29.

KERSWILL, Peter & Ann WILLIAMS (2000) "Creating a new town koine: children and language change Milton Keynes", *Language in society,* 29,1,65-116.

LABOV, William (2001) *Principles of linguistic change, Volume II, Social Factors*, Oxford: Blackwell.

LAPESA, Rafael (1981) *Historia de la lengua española*, Madrid, Gredos.

LÓPEZ-MORALES, Humberto (1980) *Estratificación social del español de San Juan de Puerto Rico*, México D.F.: Universidad Nacional Autónoma de México.

LINZ, Juan (1986) *Conflicto en Euskadi*, Madrid: Espasa-Calpe.

LIPSKI, John (1996) *El español de América*, Madrid: Cátedra.

MILROY, Lesley (1987) *Languages and Social networks,* 2.ª ed., Oxford: Blackwell.

MITXELENA, Luis (1977) *La lengua vasca*, Durango: Leopoldo Zugaza.

MORENO FERNÁNDEZ, Francisco (1996) "Castilla la Nueva", en Manuel Alvar (ed.), *Manual de dialectología hispánica*: *El español de España*, Barcelona: Ariel, 213-232.

PAYNE, Stanley (1975) *Basque Nationalism*, Reno: University of Nevada Press.

PAYNE, Arvilla (1980) "Factors controlling the acquisition of the Philadelphia dialect by out of state children", en William Labov (ed.), *Locating language in time and space*, Nueva York, Estados Unidos: Academic Press, 143-158.

POTOWSKY, Kim (2011) "Intrafamilial dialect contact", in Manuel Díaz-Campos (ed.), *Handbook of Hispanic Sociolinguistics,* Hoboken: Wiley-Blackwell, 579-597.

RUIZ OLABUÉNAGA, José Ignacio & Cristina BLANCO (1994) *La inmigración vasca. Análisis trigeneracional de 150 años de inmigración*, Bilbao: Universidad de Deusto.

SAMPER, José Antonio (1990) *Estudio sociolingüístico del español de Las Palmas de Gran Canaria,* Santa Cruz de Tenerife: La Caja de Canarias.

SECO, Manuel (1988) *Diccionario de dudas y dificultades de la lengua española*, Madrid: Espasa-Calpe.

TRUDGILL, Peter (1986) *Dialects in contact*, Oxford: Blackwell.

URRUTIA, Hernán (1988) "El español en el País Vasco: peculiaridades morfosintácticas", *Letras Deusto*, 18, 33-46

VANN, Robert (1996) *Pragmatic and cultural aspects of an emergent language variety: the construction of Catalan Spanish deictic expressions*, Tesis Doctoral en Lingüística Hispánica, Austin: Universidad de Tejas.

VILLENA-PONSODA, José Antonio (2005) "How similar are people who speak alike?", en Peter Auer, Frans Hinskens & Peter Kerswill (eds.), *Dialect Change: Convergence and Divergence in European Language*, Cambridge: Cambridge UP, 303-334.

ZAMORA VICENTE, Alonso (1967) *Dialectología española*, Madrid: Gredos.

ZÁRATE, Maite (1976) *Influencias del vascuence en la lengua castellano,* Bilbao: Editorial de la Gran Enciclopedia Vasca.

Dialectologia 17 *(2016), 51-70.*
ISSN: 2013-2247

Received 1 December 2014.
Accepted 10 March 2015.

THREE TYPES OF PREPOSITIONS IN SPANISH *SE* SENTENCES. CONSEQUENCES FOR CROSS-DIALECTAL STUDIES

Ángel J. GALLEGO

Universitat Autònoma de Barcelona*

angel.gallego@uab.cat

Abstract

This paper discusses agreement patterns of SE sentences in different Spanish dialects. Special attention is paid to situations where the verb agrees with Case-marked internal arguments (cf. Torrego 1998, López 2012) bypassing the preposition (e.g., *Se ayudaron a los banqueros*, Eng. 'Bankers were helped'), and to a previously unnoticed case in which agreement occurs across a non-clitic related preposition (e.g., *Se saben de diversos factores*, Eng. 'Different factors are known'). A micro-parametric approach is put forward whereby two functional elements hold the key to accounting for the facts: on the one hand, the feature specification of v and T (the locus of structural Case) may vary, and, on the other, the precise nature of what we label "P" may range over three possible manifestations: (i) a *bona fide* preposition, (ii) an applicative element (potentially associated to a clitic), and (iii) the spell-out of a feature within a given functional category.

Keywords

Spanish, impersonal / passive se, syntax, agreement, prepositions

* Departament de Filologia Espanyola. Facultat de Filosofia i Lletres. Edifici B. Universitat Autònoma de Barcelona. 08193 Bellaterra (Barcelona), Spain.

TRES TIPOS DE PREPOSICIONES EN ORACIONES CON *SE* DEL ESPAÑOL.
CONSECUENCIAS PARA ESTUDIOS DIALECTALES

Resumen

Este artículo discute los patrones de concordancia de oraciones con SE en diferentes dialectos del español. Se presta especial atención a situaciones en las que el verbo concuerda con argumentos internos que han recibido caso (cf. Torrego 1998, López 2012), ignorando la preposición que los introduce (e.g., *Se ayudaron a los banqueros*), y a una variante no descrita previamente en la que la concordancia tiene lugar a través de una preposición no relacionada con clíticos (e.g., *Se saben de diversos factores*). El presente trabajo ofrece un planteamiento micro-paramétrico en el que dos elementos funcionales son clave para dar cuenta de los hechos: por un lado, la especificación morfológica de v y T (el *locus* del caso estructural) puede variar, y, por el otro, la naturaleza específica de lo que llamamos "P" puede adoptar tres manifestaciones: (i) una preposición *bona fide*, (ii) un elemento aplicativo (potencialmente asociado a un clítico), y (iii) la manifestación de un rasgo de una categoría funcional.

Palabras clave

español, impersonal / pasiva con SE, sintaxis, concordancia, preposiciones

1. Introduction

It is well-known that *preposition stranding* is a cross-linguistically restricted phenomenon (cf. Law 2006 and references therein for discussion). Thus, Romance languages such as Spanish prevent instances of A-bar movement stranding a preposition, as noted by Campos (1991):

(1) *Quién contaron todos con? (Spanish)
who counted all with
Who did everybody count on?

[from Campos 1991: 741]

Whatever the factor responsible for (1) (cf. Abels 2003, Hornstein & Weinberg 1981, Kayne 1984, and Truswell 2009 for different accounts), it plausibly holds in the case of pseudopassives, which are ruled out too:

(2) *José es contado con por todos (Spanish)
José be counted with by everybody
José is counted on by everybody
[from Campos 1991: 741]

The literature on these phenomena has emphasized the empirical observation that pseudopassivization is more restricted than P-stranding (cf. Abels 2003 and Truswell 2009). The goal of this short paper is to discuss previously unnoticed data from non-standard Spanish that indicate that this language can display a pseudopassive pattern in the context of "SE passives." Interestingly, pseudopassivization is barred with "BE (or periphrastic) passives," which we take to reinforce the structural and morphological differences of the vP of SE and BE passives (cf. Mendikoetxea 1992, 1999).

The paper is divided as follows. Section 2 provides overview of the agreement options of SE sentences. Sections 3 and 4 discuss the properties of what is called "hybrid pattern" and what I call "residual pseudopassives" respectively; section 3 further outlines an account of the facts that capitalizes on the properties of functional categories, thus adopting a micro-parametric approach. Section 5 summarizes the main conclusions.

2. SE sentences: basic properties

The literature on SE sentences has discussed the morphological and syntactic intricacies associated to this clitic (cf. Raposo & Uriagereka 1996; D'Alessandro 2007; Mendikoetxea 1992, 1999; and López 2007, among others). In the case of Spanish, it is known that SE can participate in both passive (agreeing) and impersonal (non-agreeing) sentences:

(3) a. Se vendieron los coches PASSIVE SE (Spanish)
SE sold-3.pl the cars
The cars were sold

b. Se ayudó a los estudiantes IMPERSONAL SE (Spanish)
SE helped-3.sg to the students
The students were helped

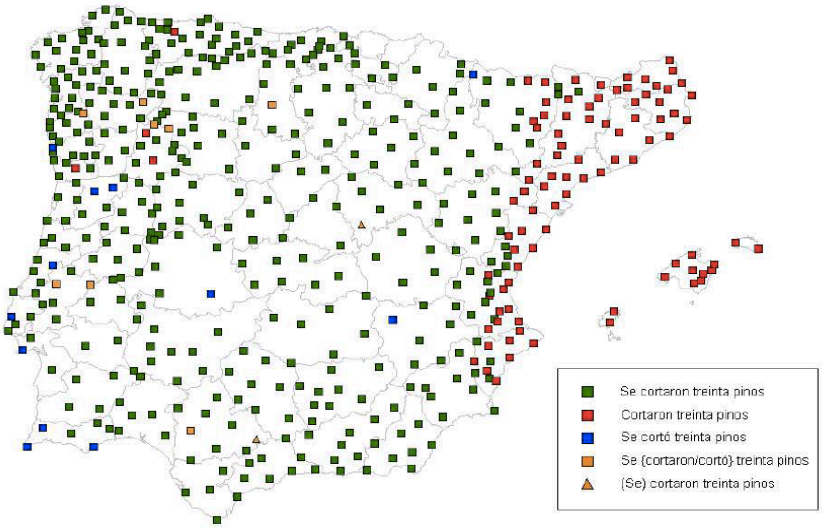
(3b) is a transivitive sentence, which in a system like Chomsky's (2001) means that v is ϕ-complete and assigns accusative Case to the DP object *los estudiantes*, SE plausibly occupying the position of the external argument (as argued by Raposo & Uriagereka 1996 and López 2007). (3a), on the other hand, is a passive structure, where v is ϕ-defective, and the internal argument receives nominative Case from T. Like in BE (or periphrastic) passives, the subject can remain in its base-generation position or move to [Spec, TP]:

(4) a. Se vendieron los coches (Spanish)
SE sold-3.pl the cars
The cars were sold

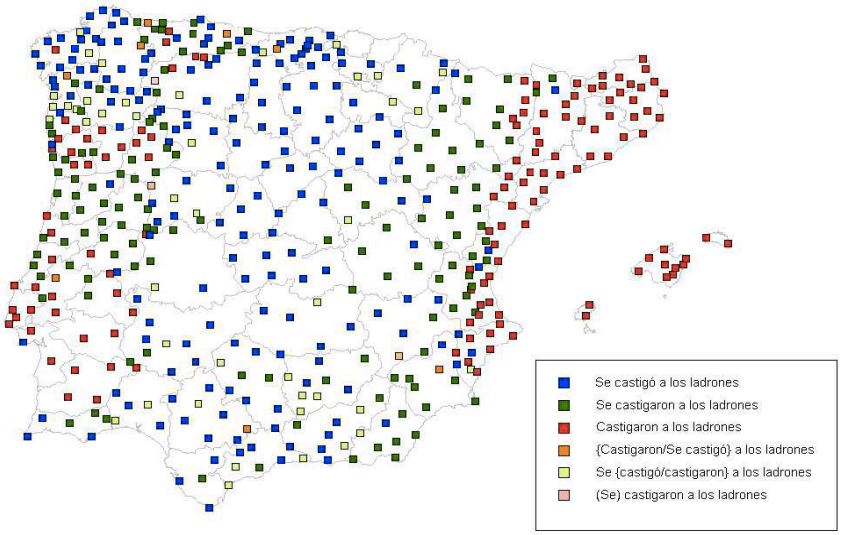
b. Los coches se vendieron (Spanish)
the cars SE sold-3.pl
The cars were sold

These two options for SE sentences have been documented in traditional atlases, like Tomás NavarroTomás' ill-fated ALPI (*Atlas Lingüístico de la Península Ibérica*):

(5) a. Se **cortaron treinta pinos** (Eng. 'Thirty pines were cut')



b. Se **castigó** a **los ladrones** (Eng. 'Thieves were punished')



[from de Benito 2010: 8, 14]

One other well-known fact is that SE passives align with BE passives in many respects. Interestingly enough, Mendikoetxea (1999: §26.3.2.2.) notes that SE passives can manifest either full (person, number) or partial (defective) agreement, a traditional observation that goes back to Bello (1847) (cf. Martín Zorraquino 1979 for additional discussion):

(6) a. Se venden botellas — PASSIVE 1 (full agreement) (Spanish)
SE sell-3.pl bottles
Bottles were sold

b. Se vende botellas — PASSIVE 2 (defective agreement) (Spanish)
SE sell-3.pl bottles
Bottles were sold

The second pattern of SE passives (non-agreeing passives, sometimes collapsed with impersonal passives) can be found already in Old Spanish, but it is also found in present-day non-European Spanish, as pointed out in Mendikoetxea (1999) and RAE-ASALE (2009). There are different factors that seem to conspire to yield the second pattern in (6) (cf. RAE-ASALE 2009). I list them below:

(7) a. The category of the internal argument (DP or NP)
b. The preverbal or postverbal position of the internal argument
c. The grammatical aspect of the verb (perfective vs. imperfective)
d. The presence of dative arguments
e. The specific proximity of the internal argument (locality conditions)

In the examples below, we can see how the just listed factors have an impact on agreement processes in SE passives (cf. RAE- ASALE 2009: §41.12c and ff.):

(8) a. Se necesita aprendices
SE need-3.sg learners
Learners are needed

a'. *?Se necesita los aprendices
SE need-3.sg the learners
Learners are needed

b. Aquí se necesita aprendices
here SE need-3.sg learners
Learners are needed here

b'. *?Aprendices se necesita aquí
learners SE need-3.sg here
Learners are needed here

c. Se vende libros
SE sell-3.sg books
Books are sold

c'. ?Se vendió libros
SE sell-3.sg books
Books were sold

d. Se les da caramelos a los niños
SE cl.dat give-3.sg candies to the children
Children are given candies

e. Se veía a un lado y a otro del camino las mansiones . . .
SE see-3.sg at one side and to other of-the track the mansions
Mansions were seen at one side and the other of the track

As for non-European varieties, RAE-ASALE (2009: 3094) notes that "The distribution is not perfect […] it has been observed that Andean, Chilean, and River Plate Spanish feature overlapping more clearly" (my translation). Some examples are given in (9), taken from RAE-ASALE (2009):

(9) a. En su partido se respeta las libertades . . . (Mexican Spanish)
in his party SE respect the freedoms
Freedoms are respected in his party

b. Se atendió once solicitudes . . . (Mexican Spanish)
SE attend eleven applications
Eleven applications were attended

To sum up so far, SE passive sentences display various agreement patterns in the different varieties of Spanish. For the most part, such patterns concern either the ϕ-complete / ϕ-defective status of T (the locus of nominative Case) or the possibility that the internal argument (the would-be subject) is within the search domain of T (cf. Chomsky 2001, Legate 2014). In any event, this variation concerns SE passives, which do not feature DOM. We would like to concentrate on SE sentences with DOM (so-called SE impersonals), for the same dichotomy is found there.

3. SE passives (1): the hybrid pattern

As noted at the outset of this paper, the clitic SE can participate in passive and impersonal structures. The relevant minimal pair was given in (3), and is repeated here as (10) for convenience:

(10) a. Se vendieron los coches — PASSIVE SE — (Spanish)
SE sold-3.pl the cars
The cars were sold

b. Se ayudó a los estudiantes — IMPERSONAL SE — (Spanish)
SE helped-3.sg to the students
The students were helped

Although the verb typically fails to agree with the internal argument in (10b), agreement does occur in some instances of Case-marked internal arguments. Abstractly, this pattern, which is dubbed "hybrid" by RAE-ASALE (2009), can be depicted as in (11):

(11) [SE **T** [$_{VP}$ V . . . [*a* **XP**]]]
│_____↑

Again, we see that agreement may or may not occur already in previous stages and in non-European varieties of Spanish:

(12) a. A **estos** no se **pueden** premiar (*Quijote*)
to these not SE can-3.pl award
These cannot be awarded

b. Se **premiaron** a **los mejores jinetes** (*Salvador Hoy*)
SE award-3.pl to the best riders
The best riders were awarded

[from RAE-ASALE 2009]

If we consider impersonal SE more closely, notice that the v of this structure should be v*, thus capable of assigning accusative. However, it seems that this Case is restricted to animate internal arguments:

(13) a. *El arroz, se **lo** come cada domingo (Spanish)
the rice SE it eat-3.sg every Sunday
The rice, it is eaten every Sunday [from Ordóñez 2004: 6]

b. A un hombre, no se **lo** juzga sin pruebas (Spanish)
to a man not SE him judge without proof
A man is not judged without evidence

This pattern seems pretty robust. So one could assume the generalization in (14):

(14) If the internal argument is Case-marked (*a*-XP), then SE v is v* (ϕ-complete)

This said, there are some exceptions. The example in (15) indicates that, in certain circumstances, v can assign accusative even with inanimate (non Case-marked) internal arguments (the sentence is adapted from Marías 2008):

(15) Cuando se reproduce lo acontecido, sin querer se **lo** deforma (Spanish)
when SE reproduce it happened without want SE it distort-3.sg
When one reproduces what has happened, one distorts it involuntarily

It seems that this pattern is highly restricted in the case of European Spanish. It is more active in non-European varieties. In particular, RAE-ASALE (2009: §41.12m) argues that accusative assigning v* with inanimate internal arguments is licensed in the Andean, Chilean, and River Plate areas.

(16) a. Se planifican los escapes, se **los** tecnologiza (Spanish)
SE plan-3.pl the escapes SE CL technologize
Escapes are planned, they are technologized

b. Fracasan solo cuando se **las** usa mal (Spanish)
fail-3.pl just when SE CL use-3.sg bad
They fail only when they are used in a wrong way

c. Se **los** entiende sin que hayan sido explicados (Spanish)
SE CL understand-3.sg without that have been explained
They are understood without having been explained

[from RAE-ASALE 2009:3098]

One more examples of this exotic pattern is (17), this time from European Spanish (cf. Martín Zorraquino 1979, Fernández-Ordóñez 1999):

(17) a. Este último [avión] ya está listo y debe ser retirado, pues por cada día que pasa y no **se lo** utiliza se pierde dinero y además hay que pagar multa
(*La Nación*, 7-IX-1975, pág. 20, c-7, apud Martín Zorraquino 1979
This last plane is ready and must be taken away, since every day that goes on and it is not used we lose money and we have to pay

b. El lomo **se lo** da una vuelta en la sartén, **se lo** mete a la olla, **se lo** cubre con aceite de oliva
The meat has to be turned upside down in the pan, you put it into the pot, you cover it with olive oil
(Campo de San Pedro, Segovia, COSER 3702, apud Fernández-Ordóñez 1999)

[from de Benito 2013: 147]

In sum, pronominalization of Case-marked internal arguments, like *a Pedro* (Eng. 'to Pedro') in (18), as in (19):

(18) Se critica **a Pedro** (Spanish)
SE criticize to Pedro
Pedro is criticized

(19) Pronominalization of (18) (# indicates that the form is not preferred)

a. Se **lo** critica (non-leísta / American Spanish)

b. Se **{#lo/le}** critica (leísta / European Spanish)

This raises the question whether Case-marked internal arguments receive true accusative. If they do not, then that would explain the restricted availability of *lo*/*la* (only with animates), and the preference for *le* in European Spanish. This process of *lo* > *le* shift with SE can be seen even by speakers that are not *leístas* with masculine in regular transitive sentences, as noted by Ordóñez (2004).

(20) Si hay que fusilar-**lo**, SE **le** fusila (European Spanish)
if there-be-3.sg that shoot-CL SE CL shoot-3.sg
If he must be shot, he is shot
[from P. Preston, *Franco*, cited by Ordóñez 2004]

Unlike European Spanish, Mexican Spanish shows no *le* clitic with standard transitive sentences — it is a non-leísta dialect. All direct objects, masculine or feminine, deploy the standard masculine vs. feminine distinction: *lo* / *la*. This can be seen in (21):

(21) a. **A** Juan **lo** vieron contento (Mexican Spanish)
to Juan CL see-3.pl happy
Juan, he was seen happy

b. **A** María **la** vieron contenta (Mexican Spanish)
to María CL see-3.pl happy
María, she was seen happy

However, in the presence of SE, Mexican Spanish obligatorily shifts to *le*.

(22) a. A Juan SE **le** vio contento (Mexican Spanish)
to Juan SE CL see happy
Juan, he was seen happy

b. A María SE **le** vio contenta (Mexican Spanish)
to María SE CL see happy
María, she was seen happy

This shift to *le* does not occur in Río de la Plata Spanish. This south-American dialect, contrary to Mexican Spanish or European Spanish, has doubling with Case-marked internal arguments beyond strong pronouns:

(23) a. (**lo**) vi a Juan (River Plate Spanish)
CL saw-1.sg to Juan
I saw Juan

b. *(**la**) vi a la libreta (River Plate Spanish)
CL saw-2.sg to the notebook
I saw the notebook

In this dialect no *le* shift occurs with direct objects:

(24) a. Se **(lo)** escuchó [al niño] (River Plate Spanish)
SE CL heard-3.sg to-the boy
The boy was heard

b. Se **(la)** escuchó [a la niña] (River Plate Spanish)
SE CL heard-3.sg to-the boy
The girl was heard

Descriptively, Spanish dialects that allow clitic doubling with Case-marked direct objects do not shift to *le* in impersonal SE constructions (cf. Ordóñez & Treviño 2007 for an account).

From all the discussion above, one can plausibly conclude that impersonal sentences with SE are divided into two dialects in Spanish:

(25) a. Dialect A: v is ϕ-defective (accusative unavailable)

b. Dialect B: v is ϕ-complete (accusative available)

Technically, this amounts to (26):

(26) a. Dialect A: $[_{vP}$ v $[_{VP}$ V $[_{PP}$ *a* [$DP_{OBLIQUE}$]] (leísta Spanish)

$\vert$___↑

b. Dialect B: i. $[_{vP}$ vϕ $[_{VP}$ V $[_{KP}$ *a* DP_{ACC}]]] (non-leísta Spanish)

$\vert$_____↑

ii. [. . . Tϕ . . . $[_{vP}$ v $[_{VP}$ V $[_{KP}$ *a* DP_{NOM}]]] (hybrid pattern)

$\vert$_____↑

As reported by Ordóñez & Treviño (2007), Mexican and Argentinian varieties of Spanish, which belong to dialect B, may show agreement with an internal argument preceded by accusative *a* (the aforementioned "hybrid pattern").

(27) a. Finalmente, se **castigaron** a **los culpables** (Mexican Spanish)
finally SE punished-3.pl to the culprits
Finally, the culprits were punished

b. Se **evacuaron** a más de **120.000 damnificados** (Argentinian Spanish)
SE evacuated-3.pl to more of 12.000 damaged
More than 120.000 damaged people were evacuated

[from Ordóñez & Treviño 2007: 12]

The data in (27) pose a puzzle. They clearly indicate that the ϕ-Probe in T can agree with the internal argument, but this is unexpected, given that the latter has already been Case marked (by v), and is thus "inactive" in Chomsky's (2001) terms. It is nonetheless possible — and it is what we would like to propose here — that dialect B divides further into a subdialect that fails to Case mark the internal argument.

A way to go about this subtler micro-parametric distinction is to take Spanish *a* to vary within the relevant varieties. Building on much literature on this topic (cf. López

2012, Torrego 1998, and references therein), we assume that the vocabulary item *a* corresponds to three different elements in Spanish:

(28) A three-way analysis for *a* in Spanish
a. A spell-out of a true preposition
b. The spell-out of a Case/clitic-related projection (cf. López 2012, Torrego 1998)
c. The spell-out of a feature of a Case/clitic-related projection

Clearly, in the varieties of Spanish that license (28b), *a* is not a preposition, and it is not the standard Case-marking morpheme of DOM — for otherwise agreement would fail —, so we are left with option (28c): *a* is the spell-out of a feature, not even a projecting category. Given that the v of dialect Bii is ϕ-defective and that *a* is not a preposition, it follows that the internal argument can long-distance agree with T.

Having considered the basic Case-agreement configurations where SE is involved, we would like to briefly consider a pattern that seems to be intimately related to the one in (28c), and which quickly evokes the profile of pseudopassive structures.

4. SE passives (2): residual pseudopassives

As just noted, the examples in (27) show that the ϕ-Probe on T can long-distance agree with the internal argument, ignoring the would-be preposition — actually a feature, under the present account — *a*. This is somewhat surprising, as it resembles a pseudopassive.

Yet much more surprisingly, other variants (mainly American) of Spanish dialect A manifest agreement with DPs contained in lexical PPs. The following data are from different on-line sources:

(29) a. Dijo que se **hablaron** con **las autoridades** (American Spanish)
say that SE talked-3.pl with the authorities
He said that the authorities were talked to
[http://www.santiagodigital.net/index.php?option=com_content&task=view&id=13837&Itemid=17]

b. En Santiago anoche se **informaron** de **cuatro homicidios** (American Spanish)
in Santiago last night SE informed-3.pl of four homicides
Four homicides were reported last night in Santiago
[http://www.periodismoglobal.cl/2006/08/la-democracia-de-la-udi.html]

c. El comercio online sumó [...] 100 millones de transacciones (American Spanish)
the trade online added-3.sg 100 millions of transactions
[...] cuando se **llegaron** a **los 74,3 millones de operaciones**
when SE arrived-3.pl to the 74,3 millions of operations
The online trading added 100 million transactions when 74,3 million operations were reached
[http://www.elpais.com/articulo/economia/comercio/electronico/volvio/batir/record/2010/elpepueco/20110506elpepueco_7/Tes]

d. En realidad se **dependen** de **tantos factores** (American Spanish)
in reality SE depend-3.pl of so-many factors
que esto provoca una extrema dificultad
that this provokes a extreme difficulty
Actually, one depends on so many factors that it makes things extremely difficult
[http://diegotenis9.wordpress.com/]

More data can be obtained from the CREA database, and from Google:

(30) a. Sólo se **disponen** de **datos de matrículas** . . . (El Salvador)
just SE dispose-3.pl of data of registration
We just have data on registration

b. Aunque no se **disponen** de **cifras exactas** . . . (Costa Rica)
although not SE dispose-3.pl of numbers exact
Although we don't have exact numbers

c. Sí se **saben** de **diversos factores** que influyen... (Spain)
yes SE know-3.pl of diverse factors that influence
We do know factors that influence
[from CREA: http://corpus.rae.es/creanet.html]

(31) a. Todavía se **confían** en los milagros (México)
yet SE trust-3.pl in the miracles
They still believe in miracles
[http://www.sinembargo.mx/30-03-2014/947521]

b. Cuando se **hablan** de **las supuestas desigualdades** (Chile)
when SE talk-3.pl of the alleged asymmetries
When they talk about the alleged asymmetries
[http://blog.lanacion.cl/2014/03/11/desigualdades-de-genero-en-el-emprendimiento/]

These data are rather restricted due to normative pressures, but they are not isolated on-line hits. The main conclusion to be drawn from (29) is that certain dialects of Spanish display, contrary to what is typically assumed, pseudopassives.

This raises at least two questions. The first one is whether, apart from "SE pseudopassives", Spanish can also display "BE pseudopassives". The answer is negative, as sentences like those in (32) are ruled out by American Spanish speakers, who find a sharp asymmetry with respect to the examples in (30-31):

(32) a. ***Fueron habladas** con **las autoridades** (American Spanish)
be-3.pl talked-3.fem.pl with the authorities
Authorites were spoken to

b. ***Fueron informados** de **cuatro homicidios** (American Spanish)
be-3.pl informed-3.masc.pl of four homicides
Four homicides were reported

The asymmetry between (30-31) and (32) provides support for the idea that SE and BE passives are morphologically and syntactically different, as has been argued in the literature (cf. Mendikoetxea 1999).

The second question is a parametric one: How does agreement take place in such varieties of Spanish? At first glance, the dialects allowing (30-31) must be able to license a 'reanalysis' process (however it must be implemented, an issue we cannot investigate here; cf. Hornstein & Weinberg 1981, Kayne 1975, 2004, among many others) whereby T can long-distance agree with the complements of P.

It is important to point out, to conclude, that even though pseudopassivization seems to be an option in Spanish, preposition stranding is still impossible. That is to say, sentences like those in (30-31) with the agreeing DP in [Spec, TP] (after A-movement) or [Spec, CP] (after A-bar movement) are impossible. What is truly surprising, and has gone unnoticed in the literature, is the very existence of the examples in (30-31). This not only suggests that Spanish does have a residual type of pseudopassives, it also seems to threaten the empirical generalization that pseudopassives are cross-linguistically more restricted than preposition stranding.

5. Conclusions

This paper has made two interesting points. On the empirical side, we have shown that, along with the hybrid pattern of SE sentences, some dialects of Spanish feature what appear to be some form of pseudopassive construction (see data in 30 and 31). Of course, a more careful study is needed, and the factors to control for are (at least) the following: (i) the type of verb (non-pronominal, agentive) that allows pseudopassives, (ii) the type of preposition that can become inert for agreement processes, (iii) the category of the agreeing element (DP or NP), and (iv) the relevant source of data (journal, newspaper, forum, CREA, Google, etc.). Quite possibly, these could just be typos or the result of oral speech, but the fact that this 'extended' hybrid (pseudopassive, if we are correct) pattern is not found with adjuncts. In other words, examples like those in (33) are unattested.

(33) a. *Se **hablaron** en **las aulas** (Spanish)
SE talk-3.pl in the class
People talk in the class

b. *Se **aspiraron** al puesto por **muchos motivos** (Spanish)
SE aspire-3.pl to-the position for many reasons
People aspire to the position for many reasons

On the theoretical side, this paper has argued that the nature of prepositions must be divided into three types. The distinction between lexical and functional (or fake) prepositions is not new in the field (cf. Abels 2003, Cuervo 2003, Demonte 1987, 1991, 1995, Pesetsky & Torrego 2004, Romero 2011), but we have tried to sharpen it in order to account for the (28b) / (28c) distinction. Much work is required in the study of functional categories, especially in the context of dialectal variation, and this paper is nothing but a small contribution to this goal.

References

ABELS, K. (2003) *Successive cyclicity, anti-locality, and adposition stranding*, PhD Dissertation, University of Connecticut.

BELLO, A. (1847) *Gramática de la lengua castellana, destinada al uso de los americanos, edición con notas de Rufino José Cuervo*, 2 vols., Ramón Trujillo (ed.), Madrid: Arco/Libros.

BENITO, C. de (2010) "Las oraciones pasivas e impersonales con SE: Estudio sobre el ALPI", *Dialectologia*, 5, 1-25. <http://www.publicacions.ub.edu/revistes/dialectologia5/>

BENITO, C. de (2013) "(Esa tela) se la descose: La pronominalización del paciente en las impersonales reflejas del español peninsular", *Borealis*, 2.2, 129-157.

CAMPOS, H. (1991) "Preposition Stranding in Spanish?", *Linguistic Inquiry*, 22, 741-750.

CHOMSKY, N. (2001) "Derivation by Phase", in M. Kenstowicz (ed.), *Ken Hale: A Life in Language*, Cambridge, MA: MIT Press, 1-52.

CUERVO, M. C. (2003) *Datives at large*, PhD dissertation, MIT.

D'ALESSANDRO, R. (2007) *Impersonal Si constructions: agreement and interpretation*, Berlin: Mouton de Gruyter.

DEMONTE, V. (1987) "C-command, prepositions and predication", *Linguistic Inquiry*, 18, 147-157.

DEMONTE, V. (1991) *Detrás de la palabra. Estudios de gramática del español*, Madrid: Alianza.

DEMONTE, V. (1995) "Dative alternation in Spanish", *Probus*, 7, 5-30.

FERNÁNDEZ-ORDÓÑEZ, I. (1999) "Leísmo, laísmo y loísmo", in I. Bosque and V. Demonte (dirs.), *Gramática Descriptiva de la Lengua Española*, Madrid, Espasa-Calpe, 1317-1397.

HORNSTEIN, N. & A. WEINBERG (1981) "Case theory and preposition stranding", *Linguistic Inquiry*, 12, 55-91.

KAYNE, R. (1975) *French syntax*, Cambridge, MA: MIT Press.

KAYNE, R. (1984) *Connectedness and binary branching*, Dordrecht: Foris.

KAYNE, R. (2004) "Prepositions as probes", in A. Belletti (ed.), *Structures and Beyond. The Cartography of Syntactic Structures*, vol. 3, Oxford, NY: Oxford University Press, 192-212.

LAW, P. (2006) "Preposition stranding", in M. Everaert & H. van Riemsdijk (eds.), *The Blackwell companion to syntax*, Oxford: Blackwell, 631-684.

LEGATE, J. (2014) *Voice and v*, Cambridge, MA: MIT Press.

LÓPEZ, L. (2007) *Locality and the Architecture of Syntactic Dependencies*, New York: Palgrave.

LÓPEZ, L. (2012) *Indefinite objects. Differential Object Marking, Scrambling and Choice Functions*, Cambridge, MA: MIT Press.

MARÍAS, J. (2008) "Sobre la dificultad de contar", *Discurso de ingreso en la Real Academia Española*, 27 de abril de 2008, Madrid: Real Academia Española.

MARTÍN ZORRAQUINO, M. A. (1979) *Las construcciones pronominales en español. Paradigma y desviaciones*, Madrid: Gredos.

MENDIKOETXEA, A. (1992) *On the nature of Agreement: The Syntax of ARB SE in Spanish*, PhD Dissertation, University of York.

MENDIKOETXEA, A. (1999) "Construcciones con *se*: medias, pasivas e impersonales", in I. Bosque and V. Demonte (eds.), *Gramática descriptiva de la lengua española*, Madrid: Espasa Calpe, 1575-1630.

ORDÓÑEZ, F. (2004) "*Se* Across Romance", Talk given at GURT conference, Georgetown University.

ORDÓÑEZ, F. & E. TREVIÑO (2007) "Unambiguous SE", Talk given at *XVII Colloquium on Generative Grammar*, Universitat de Girona (Girona), 13-15 June 2007.

PESETSKY, D. & E. TORREGO (2004) "Tense, Case, and the nature of syntactic categories", in J. Guéron & J. Lecarme (eds.), *The syntax of time*, Cambridge, MA: MIT Press, 495-537.

REAL ACADEMIA ESPAÑOLA Banco de datos (CREA) [on line]. *Corpus de referencia del español actual.* <http://www.rae.es> [Retrieved 20 April 2014].

RAE-ASALE (2009) *Nueva Gramática de la Lengua Española*, Madrid: Espasa.

RAPOSO, E. & J. URIAGEREKA (1996) "Indefinite *SE*", *Natural Language and Linguistic Theory*, 14, 749-810.

ROMERO, J. (2011) *Los dativos en español*, Madrid: Arco Libros.

TORREGO, E. (1998) *The Dependencies of Objects*, Cambridge, MA: MIT Press.

TRUSWELL, R. (2009) "Preposition-stranding, Passivisation, and Extraction from Adjuncts in Germanic", *Linguistic Variation Yearbook*, *8*, 131-177.

Dialectologia 17 *(2016), 71-89.*
ISSN: 2013-2247

Received 7 February 2016.
Accepted 18 April 2016.

A CENTURY OF LANGUAGE CHANGE IN PROGRESS.
NEW DIALECT IN TSURUOKA, JAPAN

Fumio INOUE

Tokyo University of Foreign Studies*

innowayf@nifty.com

Abstract

In this paper an original research technique developed in Japan will be introduced, and patterns and speed of geographical diffusion will be discussed on the basis of the glottogram technique. Distinction between standardization and new dialect forms will be discussed on the basis of survey results. Contour lines can be utilized to represent linguistic differences between urban and rural areas.

Keywords

new dialect, standardization, diffusion speed, apparent time, glottogram

UN SIGLO DE CAMBIO LINGÜÍSTICO EN PROCESO.
UN DIALECTO NUEVO EN TSURUOKA, JAPAN

Resumen

En este artículo se presenta una técnica original de investigación desarrollada en Japón, y se discuten los patrones y velocidad de difusión geográfica sobre la base de la técnica del glotograma. Se debate también la distinción entre estandarización y nuevas formas dialectales a partir de los resultados de la encuesta. Las líneas de contorno pueden utilizarse para representar las diferencias lingüísticas entre las zonas urbanas y rurales.

Palabras clave

nuevo dialecto, estandarización, velocidad de la difusión, tiempo aparente, glotograma

* 3-11-1 Asahicho, Fuchu, Tokyo, Japan 183-8534.

1. Shonai Glottogram

The main scenes (or stages) of this paper are in the Tsuruoka area in northern Japan. Figure 1 is a map around Tsuruoka city. The original glottogram technique was first applied in this area in Inoue (1975). The lines in the map on the right-hand side indicate the localities investigated, which are situated along a main road from south to north. The large, red arrow shows the area which will be re-analyzed in this paper. The area is shown again on the larger satellite map on the left-hand side. The blue balloon shows the site of Tsuruoka Castle (the old city center). Red circles in Tsuruoka city area show the hypothetical contour lines in and around a local city which will be discussed in section 6.2.

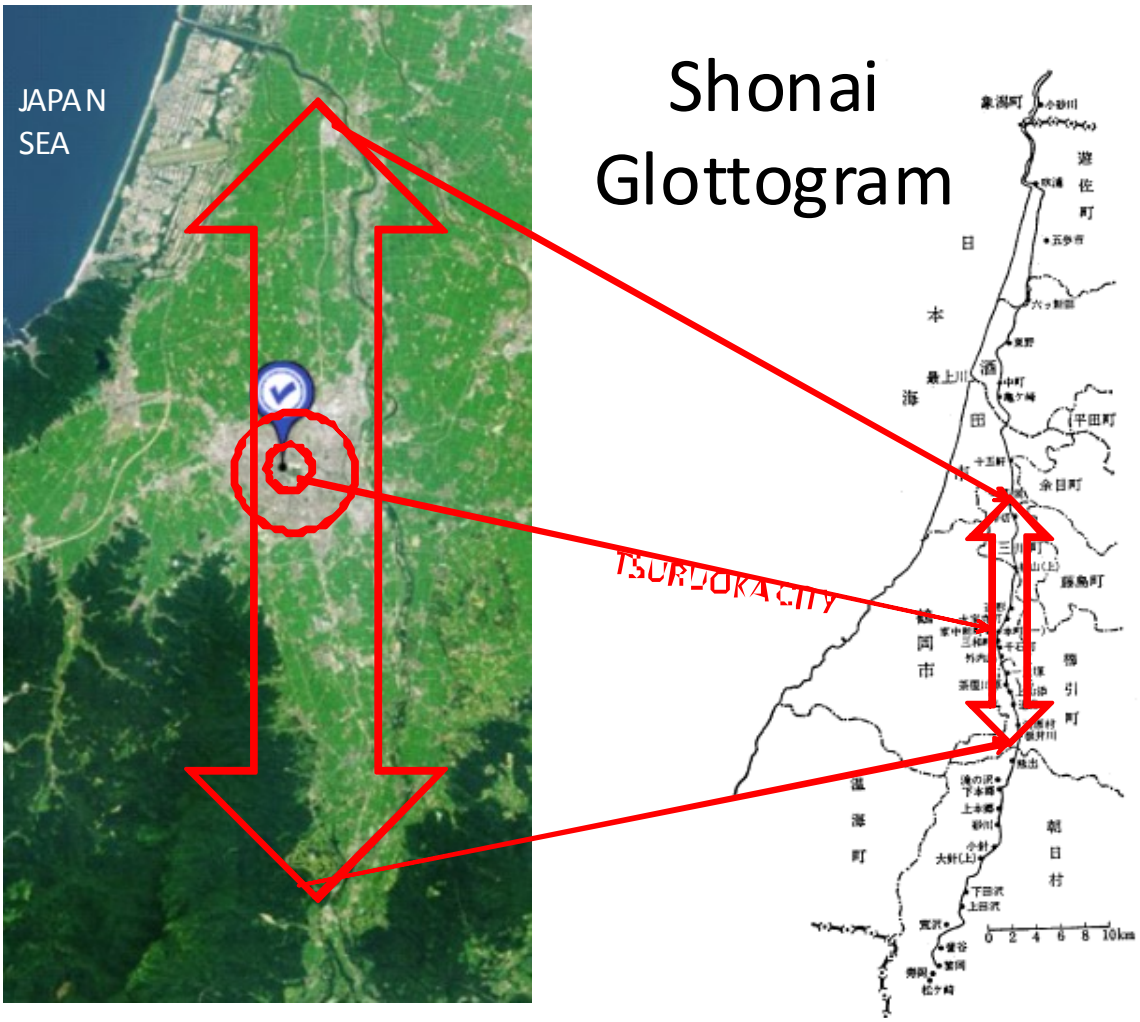


Figure 1. Localities surveyed in Tsuruoka area.

2. The Data and Methodology

2.1 Original Shonai Glottogram

Figure 2 is a sample of the original glottogram. A glottogram can be also called an "age area graph". The localities from south to north are horizontally arranged from left to right in Figure 2. Informants are arranged vertically according to age groups. Young people (teens, born around 1960) are shown below and elderly people (in their eighties, born before 1900) are shown above. The glottogram shows the apparent time differences for a time span of sixty to seventy years (Labov 1972).

The red square shows the area which will be re-analyzed in this paper. Fifteen localities are selected for re-analysis this time. These have been selected because they are situated within 12 km south and north of the Tsuruoka city area (the localities with red letters), and are dialectally under Tsuruoka's influence. The locality with the red K, the former *samurai* residential area, is postulated as the cultural center of the whole area. Localities with orange letters are suburban, and localities with green letters are rural.

The word shown in this glottogram is 'to give' or *kureru* in standard Japanese. As the purpose of this glottogram research was mainly to understand the dialect distribution patterns, standard *kureru* is shown by small dots. Dialectal form is a shortened form, *keru*. As this dialectal form is noteworthy for its increasing use by young informants, it is shown by impressive large squares, as new dialect forms.

However this glottogram is misleading in the sense that it does not faithfully show geographical distances of localities, nor the age differences of individual informants.

Shonai Glottogram (kureru > keru)

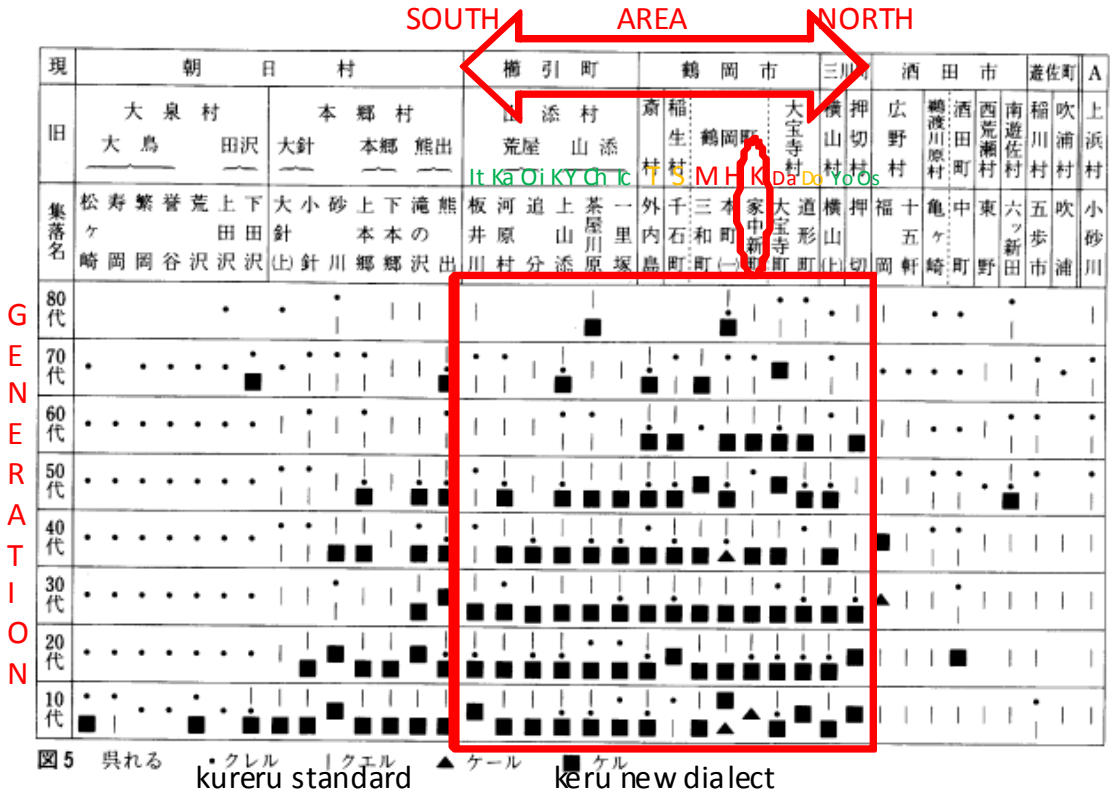


Figure 2. Original Shonai Glottogram.

2.2 Newer glottogram

In order to make a new glottogram, the "scattergram" function of EXCEL was utilized. By calculating (1) the distances of the localities from a starting point in the south and (2) the exact birth year of each informant, more faithful glottograms were drawn.

Figure 3 is a sample of a new glottogram for the standard Japanese form, *kureru*, 'to give'. Here, "2" means that only the standard form *kureru* is used. "1" means that the standard form is used together with the dialectal form, *keru*. The red, orange and green letters below show the names of the survey localities. The red ones are urban localities in the old city area of Tsuruoka. The orange are suburban rural (and industrial) localities which were consolidated in Tsuruoka City in the 20th century. The green letters show rural areas. Standard Japanese is used more in the city area.

Dialectologia 17 *(2016), 71-89.*
ISSN: 2013-2247

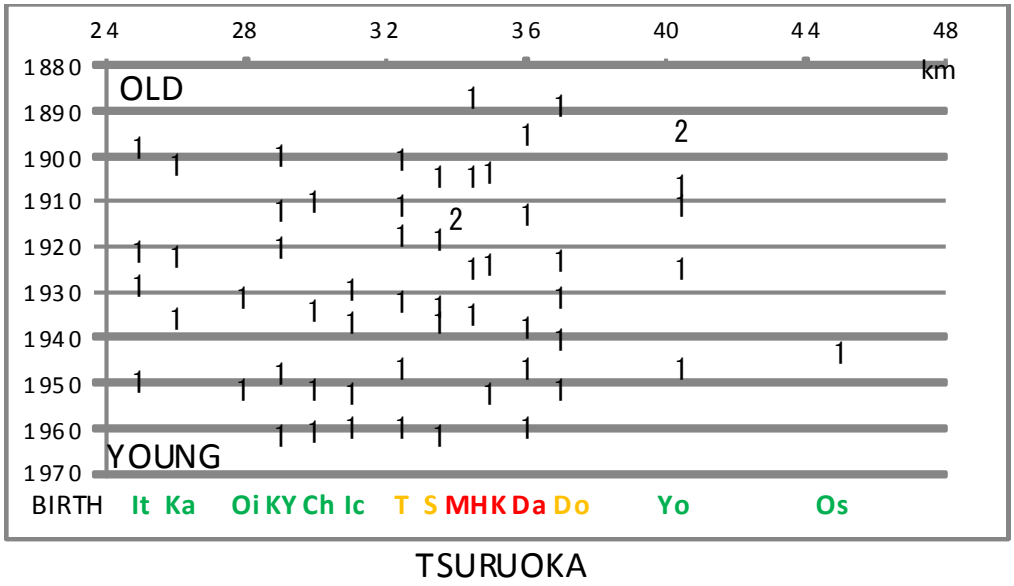


Figure 3. Glottogram of Standard Japanese *Kureru*.

Many more glottograms were drawn and they showed similar distribution patterns. Standard forms are used mainly by younger speakers near the city area.

2.3 Standardization in Shonai glottogram

In order to determine the general distribution pattern, the standard Japanese forms of 20 words have been re-analyzed as aggregate data. The total sum of usage of standard forms is shown by symbols in Figure 4**.** As expected, young speakers in the Tsuruoka city area use more standard forms than the other speakers. The blue line approximations show the general tendency. They resemble three gently-sloping mountains. The communication gap between the urban and suburban rural areas of Tsuruoka does not seem so great when compared with a new dialect form, which will be shown in Figure 6.

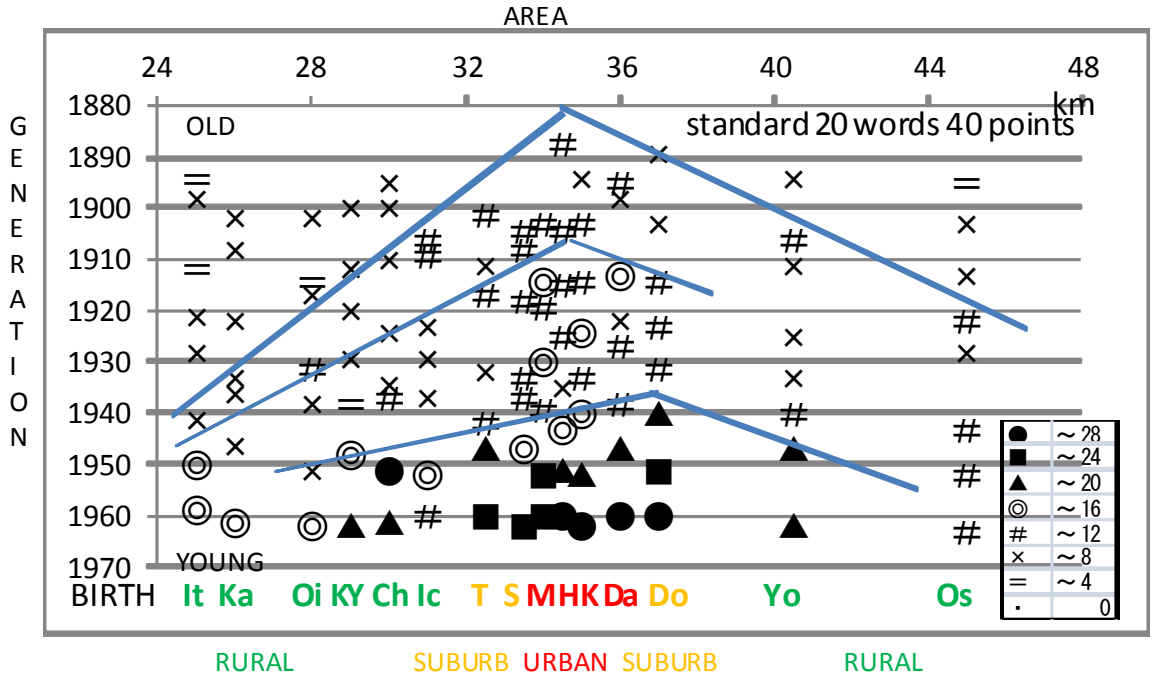


Figure 4. General distribution patterns of standard Japanese.

3. Diffusion of New Dialect in glottograms

3.1 Definition of "New Dialect"

Hereafter new dialect forms will be re-analyzed. "New dialect" is defined by three conditions: it is (1) a non-standard linguistic form, (2) used more among younger people, (3) used more in informal situations. In other words, it is a typical linguistic change in progress, and an ongoing change from below (Labov 1994, Inoue 1993, 1999).

The definition of *shin-hogen* (new dialect) in Japanese dialectology is rather different from that in English dialectology, as treated by Trudgill (1992, 2004). In Japan *shin-hogen* is used for newly-born forms, and signifies individual linguistic change. On the other hand, English new dialect according to Trudgill is a new linguistic system as a whole, such as colonial language and urban language.

3.2 Shonai Glottogram of New Dialect

Next, new dialect forms will be presented in the Shonai glottogram. This upper glottogram is the same as Figure 2. In Figure 5, the new dialect form *keru*, shown by large black squares in Figure 2, will be re-analyzed. As the distances and ages are shown faithfully, the overall distribution pattern looks like a pyramid or mountain centering on the Tsuruoka city area.

Other new dialect forms will be introduced without figures here. Standard *Omoshiro katta* ('was interesting') is expressed as *Omoshe kke* by young speakers around Tsuruoka City. However, the distribution is sporadic. Another new dialect form is *kaeda*, a shortened form of *kareda*, meaning, 'withered'. This form is used very scarcely. The sporadic use of *kaeda* among young speakers may be a revival of the old sound change of r-deletion.

When all the data of the 16 new dialect forms were assembled and shown in one glottogram in Figure 6, a pattern which appeared similar to standardization (Figure 4) emerged. That is, the use of new dialect forms is prevalent among young speakers in the city area. However a noticeable difference can be detected. The new dialect forms do not show the simple mountain type distribution. Rather, they show a pattern resembling a steep *conide* volcano, like Mt Fuji. This means that the new dialect forms born and adopted in the Tsuruoka city area are passed on or handed over to the next generation, but they do not easily diffuse or propagate to the neighboring rural areas. There seems to be a barrier between the city area and rural areas.

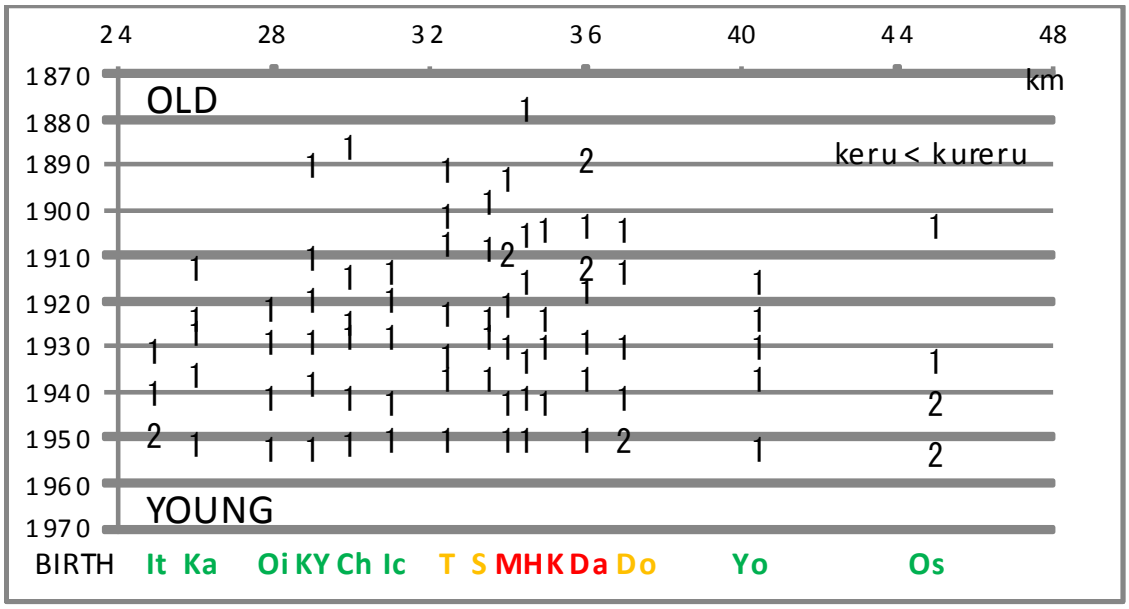


Figure 5. Glottogram of new dialect *keru*.

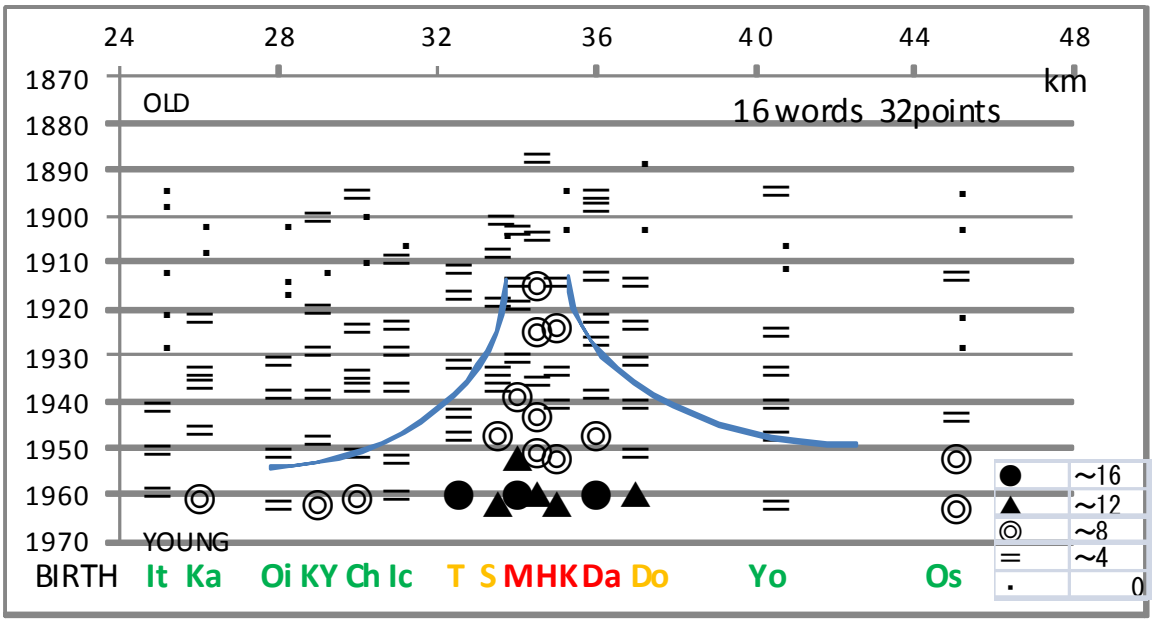


Figure 6. General distribution patterns of new dialect.

4. Geographical differences of Glottogram data

In order to grasp the general distribution pattern in the glottograms, for each locality the total usage of standard forms (Figure 4) and of new dialect forms (Figure 6)

were calculated and shown in Figure 7. Again, geographical distances are shown faithfully. Geographical differences are great as a whole. The informants from the seven (4 red and 3 orange) localities of the Tsuruoka city area are by far the heaviest users of standard Japanese and new dialect. The informants from the eight (in green, 6 southern and 2 northern) localities outside the city area do not use standard Japanese or new dialect so much.

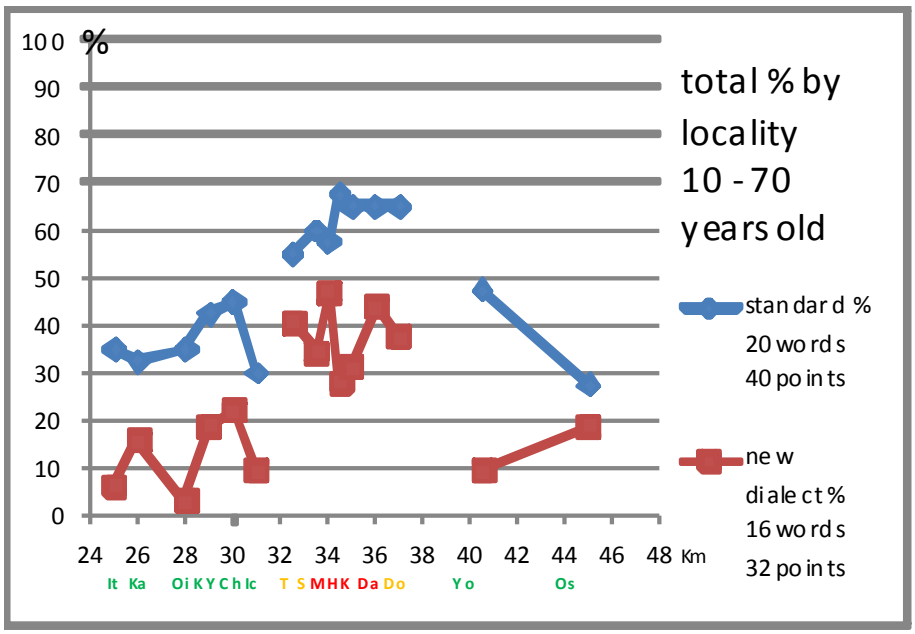


Figure 7. Distribution patterns of standard Japanese and new dialect.

4.1 Folding at the city center for standard and new dialect

The glottogram survey is usually executed along a direct line (along a road or along a railway). If we interpret urbanized area to be the center of propagation of new forms, we can metaphorically fold the glottogram in half at the city center, shifting the two peripheral rural areas to a parallel position. Figure 8 is the result of such a procedure. The old *samurai* area near the Tsuruoka castle (K) was postulated as the cultural city center. In present day, this is a high end residential area where the intellectual and white collar classes live. This location was set as 0 km and the distances to the other

locations were recalculated. The eight (6 southern and 2 northern) rural localities outside the city area were again plotted faithfully by geographical distance.

The overall tendency was simulated by approximation curves. The blue standard form is most frequently used in the locality K near the castle. In contrast, red new dialect forms are not used so much in K, while commercial and suburban industrial areas show greater usage. This corresponds with the theory that new dialect is a change from below (typically adopted by young working class people) while standardization is a change from above (typically adopted by young intellectuals).

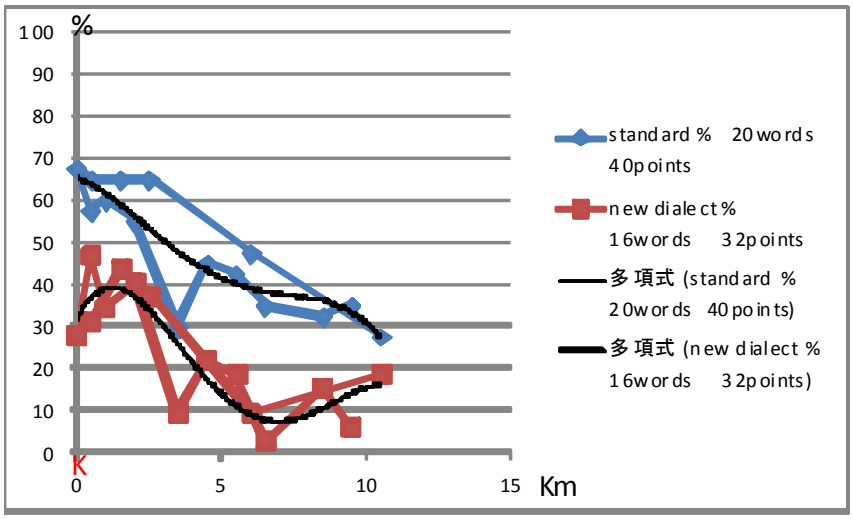


Figure 8. Folding in half at the city center for standard Japanese and new dialect.

4.2 Area Differences

In order to further simplify the overall tendency, the seven localities in the Tsuruoka city area are grouped into “urban”, and the eight localities outside the city area are grouped into “rural”. The differences in Figure 9 are great. The difference in red new dialect is especially marked, at nearly 3 times greater, between 12.9 and 37.5, whereas the difference in blue standard language is less than twice, between 36.5 and 62.1.

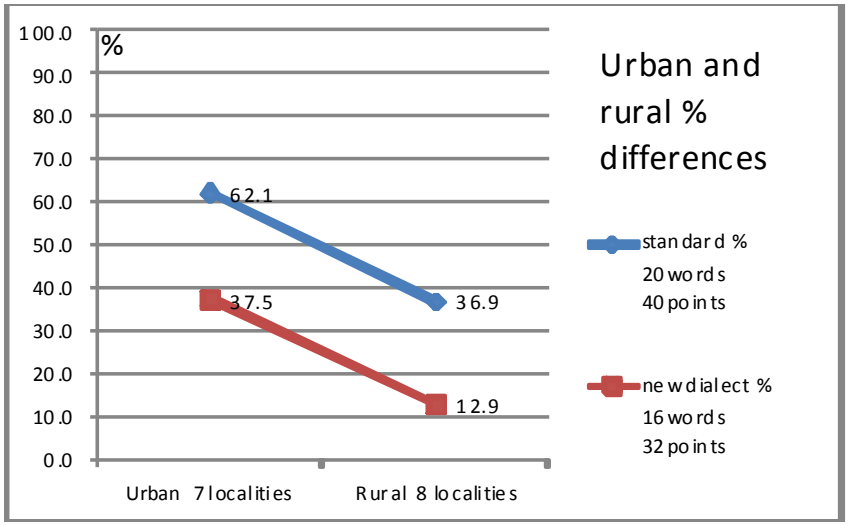


Figure 9. Urban and rural differences.

4.3 Age Differences

So far, geographical (or social) differences have been analyzed. In order to grasp the age-based differences in distribution in the glottograms, the sums of usage for standard and new dialect forms were calculated for age groups. The age differences in Figure 10 are conspicuous, especially in the case of standard Japanese. However the age-based differences in new dialect are in a sense far greater, because the relative ratio of older to younger informants is nearly ten times. New dialect forms seem to be regulated by the dimensions of both age and geography.

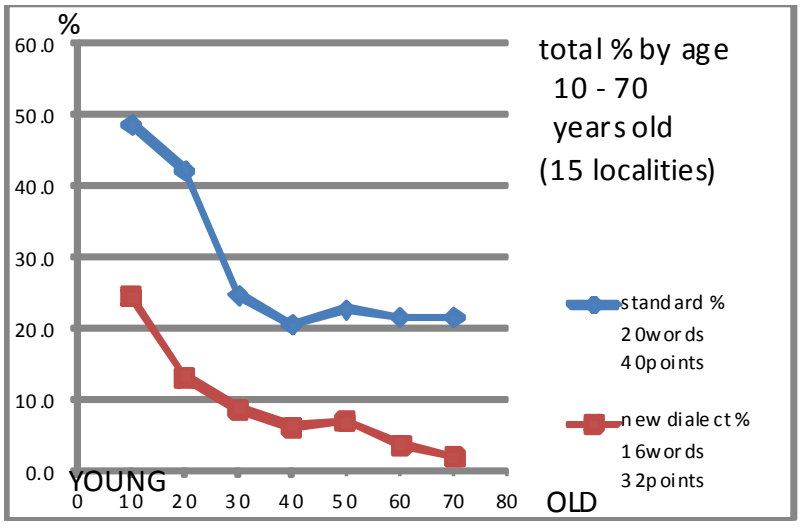


Figure 10. Age-based differences in Glottogram.

4.4 Age and Area Differences

In order to analyze the degrees of age and geographical factors at the same time, informants were further divided into four groups using the two factors. Figure 11 shows that the age factor (left two young and right two old) weighs heavier than the geographical (urban and rural) factor. Age difference is especially great in the case of new dialect forms, with a value difference of "old rural", 0.7, and "young urban", 7.3.

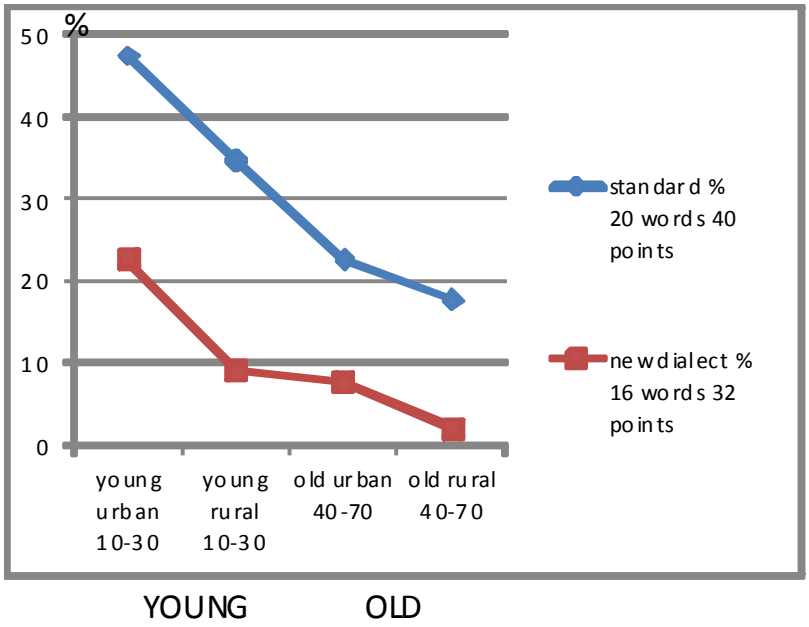


Figure 11. Age and area differences in Glottogram.

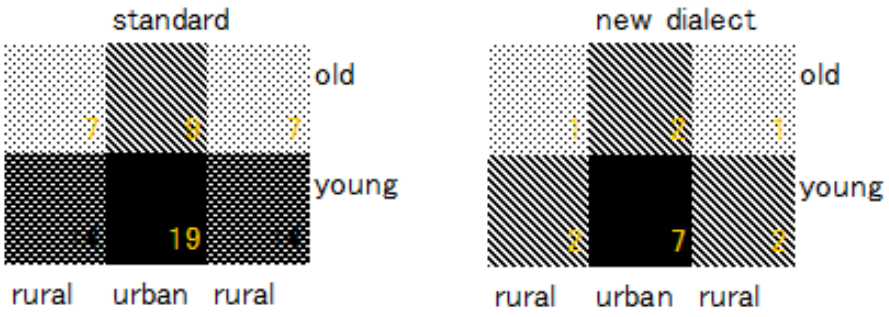


Figure 12. Age and area differences in Glottogram visualized.

The relation above is graphically (or impressionistically) represented in Figure 12 using four (or three) gradients of shade. For standard language (left) the age difference

is greater, and for new dialect forms (right) the geographical difference is also great. The left-hand figure shows that young rural people use more standard forms than elderly urban people. Meanwhile, the right-hand figure shows that the degree of usage of new dialect is similar for elderly urban people and young rural people, and that the numerical value of urban young people is strikingly great.

5. Speed of diffusion of standard and new dialect forms

5.1 Two patterns of diffusion

Thus we have seen the differences between two linguistic changes in progress, namely standardization and new dialect. Standard forms can be metaphorically envisioned as soft *lava* spreading quickly and forming a gently-sloping mountain, like an *aspite* volcano such as Hawaiian Mauna Kea. New dialect forms can be envisioned as hard *lava*, spreading slowly and forming a high and steep mountain, like a *conide*, such as Mt. Fuji.

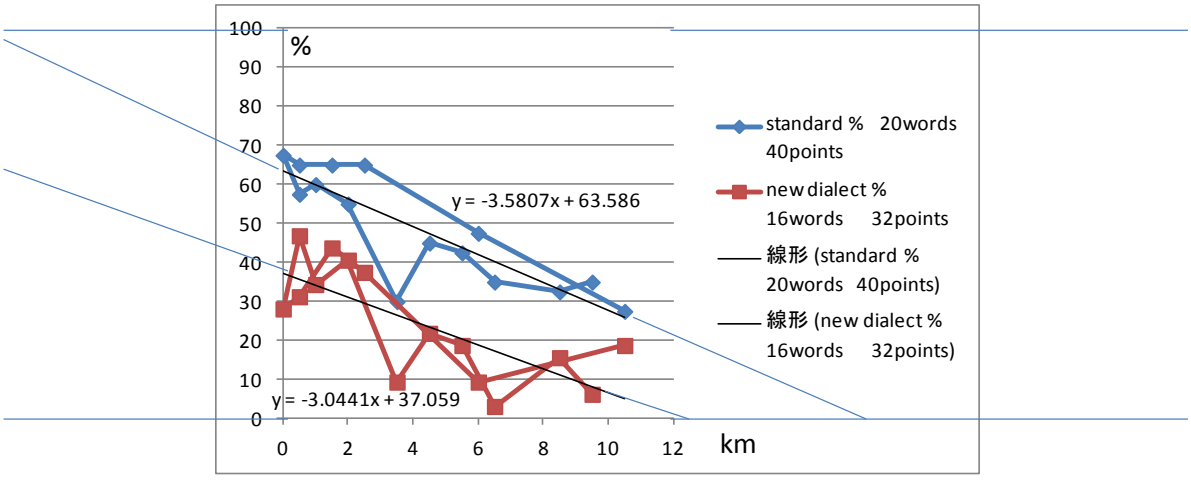
The distinct patterns of geographical diffusion of standard language and new dialect can be further distinguished using war as a metaphor. Standardization often progresses like paratroopers occupying a small (urban) spot while skipping large (rural) territories. New dialect forms progress slowly and steadily like infantry troops, occupying neighboring (rural) localities one by one.

Standardization is typically a phenomenon of modern times. In contrast, the birth and diffusion of new dialect is a process which has occurred throughout the long history of human language. We can logically surmise that similar processes of diffusion would have occurred even before modern times, and over most of the earth's surface as a universal phenomenon. New dialect is thus a kind of observatory of perennial linguistic changes.

5.2 Diffusion speed per year

On the basis of the findings above, the speed of diffusion of standard Japanese and new dialect forms can be calculated. If we apply linear approximation to the data of Fig 8, diffusion speed per year can be calculated. From the formula, shown in Figure 13, diffusion speed per year is about 0.28km/y for standard language, and 0.33 km/y for new dialect.

Standard language y=0%, x=17.7km y=100%, x=-10.2km 1÷3.5807= 0.28 km/y



New dialect y=0%, x=12.3km y=100%, x=-20. 7 km 1÷3.0441= 0.33 km/y

Figure 13. Diffusion speed on the basis of folded Glottogram.

Speed of dialect diffusion had already been calculated by Tokugawa (1993) for western dialectal forms of LAJ (*Linguistic Atlas of Japan*). Later I tried to retest the hypothesis using standard Japanese data (Inoue 2010) for 82 standard Japanese forms of LAJ. Multivariate analyses have been applied and West & East, Classical & Modern clusters were found (Inoue 2004). I also gathered glottogram data of new dialects all over Japan (Inoue 2003). In order to provide a simple figure to remember, the speed was advocated as about 1km per year. More attempts were made recently and it was found there are more varieties and possibilities. Kumagai (2013) is related to this attempt. However, speed of diffusion has not been calculated for the Kumagai data (LAJ) or the Yarimizu (2011) data of GAJ (*Grammar Atlas of Japanese Dialects*).

The variety of diffusion speed is shown in the map of Japan which was presented in Canada (Inoue 2010). Speed is shown by like a car speedometer with various angles. The speed is greater on the Pacific Ocean side, which is more densely inhabited and economically developed. It is slower on the Japan Sea side, which is sparsely inhabited and economically less-developed.

Thus far I have exaggerated differences between new dialect and standardization. There is a basic similarity between new dialect and standardization: they peak in young urban people. It is nothing new and is almost universal. Then it is reasonable to formulate this phenomenon in some quantitative way. The speeds of diffusion of standard and new dialect forms around Tsuruoka city can be calculated by making use of the volcano pattern seen in the glottograms earlier. Using age differences and locality distances, as shown in Figure 13, the average speed of diffusion was found to be around 0.3 km/y, which is slow compared to the nationwide information above. The numerical values of diffusion speed found in Shonai glottogram are low, perhaps because it represents a compact area in the Japan Seaside. This can also be explained by the somewhat scarce communication between the inhabitants of the Tsuruoka city area and the inhabitants of the rural area. The differences can be represented on a map using contour lines. High mountains and large rivers have been pointed out as obstacles to communication and propagation. These are natural obstacles. The social boundaries between urban and rural areas should be admitted as obstacles too. These hindrances to communication can be represented graphically by hypothetical contour lines as shown in Figure 1. This is the basic proposal of this paper.

5.3 Geographical and Social distances

The relation between social class and language can be represented in a graph. Urban and rural differences are basically a social distinction which still reflects the distinction from the feudal era (before Meiji Restoration in 1868). Social class differences are also reflected in expected linguistic competence, especially in the use of dialect and standard, and also in the use of honorific language, written language, and English language.

In order to calculate the diffusion speed of linguistic forms it is advisable to take into account the social dimension. The physical distance between the city area and rural areas is short, but if the social distance is taken into account, the communication distance is large. Hypothetical contour lines can be drawn around a city area to represent this social distance. The attempt to find differences between the Tsuruoka city area and the suburbs so far has shown that the differences amount to approximately one generation's gap (Inoue 2010), which is in congruence with the results of this re-analysis of the glottogram data.

6. Double Umbrella Model

6.1 Umbrella Model

An umbrella model was once constructed (Inoue 1993, 1999, 2008) to explain the mechanism of linguistic changes from below and from above based on the standard and new dialect forms. Standardization works as a pressure from above to all the areas of Japan. New dialect forms are created and adopted independently in many places in Japan. At the colloquial level, daily speech in Tokyo has the same status as other local dialects. Situated at the rim of an umbrella, Tokyoites and local people exchange new dialectal words and expressions.

The above discussions have shown that local cities like Tsuruoka have their own prestige or power of diffusion. This suggests that small umbrellas should be added on various places as is shown with small umbrellas by this revised graph in Figure 14. This can be called a “double umbrella model”.

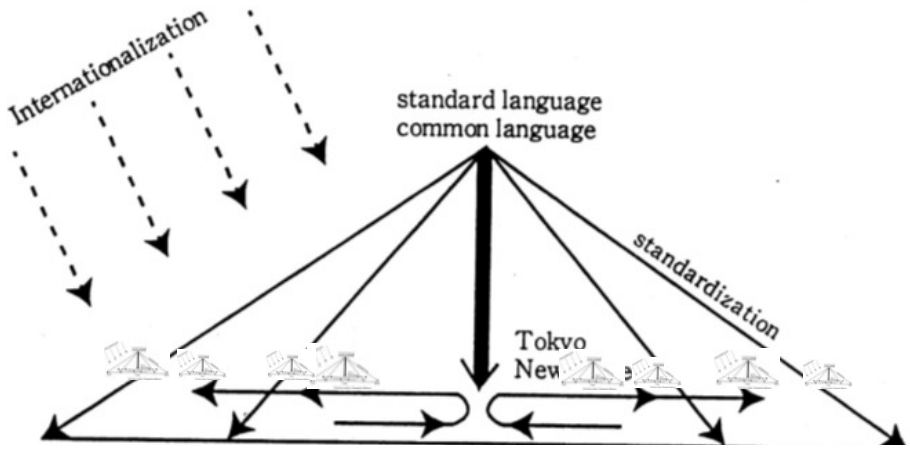


Figure 14. Double umbrella model.

6.2 Geographical and Social dialects

Thus far in dialectology NORM or "Non-mobile Older Rural Male" has been the main target of the study (Chambers & Trudgill 1980). Sociolinguistic studies of urban areas have undergone development since the latter half of the 20th century, and are mainly concerned with the various socioeconomic strata. As the other contrastive extreme of NORM, study of MYUF or "Mobile Young Urban Female" may be crucial in order to properly explain the patterns of linguistic change of standardization and new dialect. The comparison of the inner structure of local cities and their surrounding rural areas, as discussed in this paper, is one step toward the unification of the two main research fields of linguistic geography and sociolinguistics.

In conclusion, we can call the phenomena discussed above, "rural lag". This is a mirror image of "urban prestige". This is a commonsense of dialectologists and sociolinguists. By putting this usual, common process as a basic principle, the exceptional behavior of some new dialect forms which were created in rural areas and diffuse to urban areas will become more prominent. A number of concrete examples of reverse (backward) trends of some "Tokyo new dialect" have been discovered over the past several decades (Inoue 2003, 2010). However, in order to actually measure social and geographical differences, this phenomenon must be studied more formally. These

class differences between urban and rural inhabitants can be represented by mountains in urban areas using contours on the map as was shown in Figure 1.

7. Conclusions

To conclude, a double umbrella model was advocated in this paper. Local cities can be represented as small mountains on a map. They have their own power of propagation and a certain amount of labor is necessary to enter and clime inside. The labor can be represented on a map with contour lines. Rural lag should be taken into account to calculate the speed of diffusion. The speed of diffusion can be calculated more realistically if we take account of the labor of approaching the top of this urban mountain. This way of reasoning will lead to a (happy and) fruitful unification of dialectology and sociolinguistics.

References

CHAMBERS, J. K. & P. Trudgill (1980) *Dialectology,* Cambridge: Cambridge University Press.

INOUE, Fumio* (1975) "Chiriteki denpa no chosa kibo" (The scale of linguistic geographical surveys)", *Kokugogaku*, 101,17-28. (Accessible on the internet.)

INOUE, Fumio (1993) "The significance of new dialects" *Dialectologia et Geolinguistica*, 1, 3-27.

INOUE, Fumio (1999) "Trends of linguistic variations in modern Japan", *Area and Culture Studies*, 58, 39-46. <http://dictionary.sanseido-publ.co.jp/affil/person/inoue_fumio/doc/E28.pdf>.

INOUE, Fumio (2003) *Nihongo wa Nensoku 1 kiro de Ugoku (The Japanese Language Moves at 1 km per year)*, Tokyo: Kodansha Gendai Shinsho.

INOUE, Fumio (2004) "Multivariate analysis, geographical gravity centers and the history of the standard Japanese forms", *Area and Culture Studies*, 68, 15-36. (Accessible on the internet.)

INOUE, Fumio (2008) "Geographical distance center and multivariate analysis of the standard Japanese", *Dialectologia*, 1, 65-81.
<http://www.publicacions.ub.es/revistes/dialectologia1/>.

INOUE, Fumio (2010) "Real and apparent time clues to the speed of dialect diffusion", *Dialectologia*, 5, 45-64. <http://www.publicacions.ub.edu/revistes/dialectologia5/>.

KUMAGAI, Yasuo (2013) "Development of a way to visualize and observe linguistic similarities on a linguistic atlas", <http://pj.ninjal.ac.jp/socioling/nwavap02/Kumagai-NWAVAP2-2013.pdf>.

LABOV, William (1972) *Sociolinguistic Patterns*, Philadelphia: University of Pennsylvania Press.

LABOV, William (1994) *Principles of Linguistic Change*. *Internal Factors,* vol. 1, Oxford: Blackwell.

TOKUGAWA, Munemasa (1993) *Hogen chirigaku no tenkai* [Development of dialect geography], Tokyo: Hituzi Syobo.

TRUDGILL, Peter (1992) *Introducing language and society*, London: Penguin Books.

TRUDGILL, Peter (2004) *New-Dialect Formation: The Inevitability of Colonial Englishes*, Edinburgh: Edinburgh University Press.

YARIMIZU, Kanetaka (2011) "Language Change from the Viewpoint of Distribution Patterns of Standard Japanese Forms", Yuji Kawaguchi, Makoto Minegishi & Wolfgang Viereck (eds.), *Corpus-based Analysis and Diachronic Linguistics*, Amsterdam: John Benjamins, 265-284.

*Most of the English papers by F. Inoue are accessible through the internet.

http://innowayf.net/

http://dictionary.sanseido-publ.co.jp/affil/person/inoue_fumio/doc/

http://www.urayasu.meikai.ac.jp/japanese/meikainihongo/18ex/achievements.xls

Dialectologia 17 *(2016), 91-109.*
ISSN: 2013-2247

Received 17 September 2014.
Accepted 11 January 2015.

LANGUAGE PERSONALITY AS A NEW OBJECT FOR DIALECTOLOGICAL RESEARCH[1]

Ekaterina IVANTSOVA

National Research Tomsk State University*

ekivancova@yandex.ru

Abstract

The article is devoted to a new object of dialectology – a language personality. This is the phenomenon of specific social and personal traits of an individual native speaker being reflected in the text the speaker creates. It analyzes the research that arose at the junction of traditional dialectology and the theory of lingvopersonology that is being performed today by Russian dialectologists. The author examines the main projects that study the speech of an individual dialect speaker, typical features of the individuals under research, types of sources used by scientists, classical and new methods of collecting and analyzing speech material, and aspects of research of individual speech of representatives of national dialects. Prospects of this research for dialectology and other fields of the science of language are identified.

Keywords

local dialects, language personality, individual speech, Russian dialectology, lingvopersonology

[1] The study was performed by a grant of Russian Science Foundation (project № 16-18-02043).

* Томск, 634050, пр. Ленина, 36. Department of Russian Language, National Research Tomsk State University, Lenin Avenue 36, Tomsk, Russia; postcode 634050

LA PERSONALIDAD LINGÜÍSTICA COM NUEVO OBJETO DE INVESTIGACIÓN DIALECTAL

Resumen

Este artículo está dedicado a un nuevo objeto de la dialectología: la personalidad lingüística. Este es el fenómeno de rasgos sociales y personales específicos de un hablante nativo individual que se refleja en el texto que el hablante crea. Se analiza la investigación que surgió en el cruce entre la dialectología tradicional y la teoría de la linguopersonalidad que realizan actualmente los dialectólogos rusos. La autora examina los principales proyectos que estudian el habla de un hablante individual de un dialecto, las características típicas de los individuos bajo investigación, los tipos de fuentes utilizadas por los científicos, los métodos clásicos y nuevos de recolección y análisis de material hablado, y aspectos de la investigación de rasgos representativos de habla individual de dialectos nacionales. Se identifican las perspectivas de esta investigación para la dialectología y otros campos de la ciencia del lenguaje.

Palabras clave

dialectos locales, personalidad lingüística, habla individual, dialectología rusa, linguopersonalidad

1. Introduction

Linguistics has come a long way in the past two centuries, each step bringing it closer to understanding that a person is the main object of scientific knowledge. It was in the nineteenth century that Buschmann & Humboldt (2000), Osthoff & Brugmann (1881), Paul (2002), Baudouin de Courtenay (1963) and others raised the question of a ratio of the general and the individual in a language. At the beginning of the twentieth century special attention was paid to theoretical problems of studying individual speech in works of Sapir (1949), Bakhtin (1996) and Vinogradov (1980). Weisgerber (1929) defined the individual usage of a language system as one of the forms of existence of a language. In the middle of the century foundations of a communicative, functional approach, and a little later of a cognitive and linguo-culturological one, were laid in linguistics, and supplemented the traditional historical and the system-structural analyses. The central paradigm of scientific knowledge came to be anthropocentric in many sciences, including linguistics. More and more often properties of a language were analyzed for the purpose of revealing intrinsic characteristics of a person as its bearer, and vice versa: the speech of certain members of a linguistic community is a source that helps to understand properties of the language system. At last, at the beginning of the

twenty-first century, lingvopersonology became an independent discipline with its object, methods, terminology, and a base of sources (see more: Ivantsova 2010). The task of the new area of linguistic science is to study the phenomenon of the language personality. The language personality denotes a concrete, real-life native speaker whose personal characteristics synthesizing his or her social and individual traits are reflected in the texts the speaker creates.

"Identity linguistics" made headway in the work of scientists from different countries (Pound 1947; Petkov 1983; Fillmore, Kempler & Wang 1979; Johnstone 1996; Asahi 2009, and others), especially in the sphere of the so-called author's lexicography (dictionaries of those who founded national languages – Shakespeare, Milton, Pushkin, Goethe, Schiller and many others are widely known). At the same time, the theory and practice of linguistic personological research is developing more actively in Russian linguistics (Bogin 1984; Karaulov 1987; Neroznak 2003; Golev 2004, and others), where a number of centers for studying language personality are being established, in Moscow, Perm, Tomsk, Saratov, Krasnoyarsk, Blagoveshchensk, and elsewhere.

Dialectology plays an important role in the formation of lingvopersonology. The traditional object of this sphere is the speech of a community that speaks this or that locally limited subsystem of a nationwide language. However, due to the fact that dialects lack a written form, dialectologists had always focused on collecting language material in direct contact with dialects' speakers, taking into account some specific features of informants. Problems of dialectological research also facilitated raising and solving questions about a ratio of speech manifestations of an individual dialect speaker and the speaker's native dialect as a usual system, and about mental features and social and cultural particularities of a dialect community and individuals who belong to it. Now a dialect language personality has emerged as one of the central objects of linguistic research.

The purpose of this article is to analyze articulation of the linguistic personological direction in dialectological research and to characterize achievements of Russian linguistics in this area.

2. Attention to the speech of individual dialect speakers

Analysis of dialectological research of the last century shows that interest in the personality of an individual dialect speaker has constantly increased. The main projects in this area are listed below:

• 1914. An outstanding Russian philologist and historian A. Shakhmatov described one of the dialects of Ryazan Province, relying on data of a single speaker of this dialect, I.S. Grishkin (Shakhmatov 1914).

• 1949-2003. V. Timofeyev systematically studied the speech of E. M. Timofeyeva, a native of the Kurgan Region born in 1897 (Timofeyev 1971, 2003).[2]

• From 1963 to the present. A group of scientists of Perm University has been investigating the speech of A. G. Gorshkova, an inhabitant of the Perm region, born in 1891 (Gruzberg & Egoryeva 1969; Skitova & Ogiyenko 1971; Malysheva 2007).

• 1971-2005. V. Lyutikova analyzed the speech of a dialect speaker V. M. Petukhova, born in 1920, from the Kurgan region (Lyutikova 1999, 2000).

• From 1981 to the present. Linguists of Tomsk dialectological school have been carrying out research of the personality of an elderly resident of Siberia V. P. Vershinina, born in 1909 (Gyngazova 2001, 2008, 2009, 2010; Ivantsova 2002, 2005, 2009, 2006-2012, 2014; Volkova 2004; Kazakova 2007; Kuznetsova 2015, etc.).

• From 1984 to the present. The speech of members of the Lykov family, Old Believers living in the wild Sayan taiga for several decades in isolation from the outside world because of their religious beliefs, has become an object of linguistic analysis (Almukhamedova *et al.* 1986; Slesareva 1997; Markelov 2000; Tolstova 2004, 2007).

• From 1987 to the present time. Perm linguists headed by I. Russinova have been studying the personality of M. P. Suslova, born in 1926, an inhabitant of the Perm Region (Russinova 2007).

• 1990-2008. E. Nefyodova undertook research of the speech of A. I. Ponomareva, born in 1928 in the Arkhangelsk Region (Nefyodova 1997, 2000, 2001).

[2] Because the volume of magazine publications is limited, hereinafter I refer only to the most significant works of the authors or to monographic studies, which summarize the results of the preceding articles.

• 2002-2012. E. Prokofieva studied the speech of A. V. Medvedeva, a dialect language personality from the Altai Krai, born in 1913 (Prokofieva 2012).

• At the turn of the twenty-first century, a series of speech portraits of dialect speakers was created in different dialectological centers (Oglezneva 2004; Kasatkin 2007; Baklanova 2008; Volkova & Safonova 2010) and voluminous records of oral and written texts of individual representatives of folk speech culture were published (Ossipov 1995; Russinova 2007; Felde 2010; Batyreva 2011).

As one may see, the number of objects of lingvopersonological research has steadily increased. The number of large-scale projects carried out by research teams is growing. Practically all of them (except for Shakhmatov's (1914) first experience) are long-term, over ten and more years. Study of a number of dialect language personalities that began several decades ago is still underway.

3. Objects of research

Individual speakers of local dialects are typical objects of dialectological research. These are mainly elderly women who were born at the end of the nineteenth and the beginning of the twentieth century. They are mostly semiliterate members of large peasant families, and engaged in unskilled physical work all their lives. Having been raised in the environment characteristic of the Russian peasantry, they retained features of language and mentality that are particular to traditional Russian culture. The speech of these village inhabitants, though bearing common typological features, differs in its expressiveness and manifestation of the source of individuality. The language of generations of young and middle-aged dialect speakers, who are subjected to strong influence of the literary language, has not been analyzed yet from the lingvopersonological perspective.

Informants represent the main dialects of the Russian language and different dialect zones and regions. Scientists have given special attention to speakers of the North Russian dialect (the South Russian is represented much more weakly) and the

dialects of secondary formations, created in the territories of later settlements (Siberia and the Far East).

4. Data sources

First of all, records of individual speech of dialect speakers serve as sources for studying their oral speech. Until the 1970s they were handwritten without the use of technology. The majority of texts of the later period were recorded on a magnetic tape. Materials gathered in recent years are partly represented on digital media.

It is difficult to correlate data on the volume of records made as they are not present in every publication. When they are mentioned, the volume varies considerably and is measured in different units: more than 200 pages computer-typed, 1400 expressive units (Nefyodova 2000), 2476 pages of records made by hand (Gruzberg & Egoryeva 1969), 16 hours of tape recordings (Almukhamedova *et al.* 1986), and over 5000 tests (Lyutikova 1999). The biggest idiolect data archive – about 10,000 pages of records of speech of a Russian longtime resident of Siberia transcribed from a tape recorder – belongs to Tomsk linguists (Ivantsova 2006-2012, vol. 1, 13).

The range of sources to study the dialect language personality is increased by written texts, which in general are less typical for the unwritten folk speech culture. So, when studying Agafya Lykova's language personality, about 100 letters written to different addressees using half-uncial writing are used (Tolstova 2004). Autobiographical notes of a Siberian peasant V. A. Plotnikov (Ossipov 1995) and the unusual diaries of a dialect speaker M. P. Suslova have been published and described (Russinova 2007).

5. Methods of collecting material

Methods of collecting material are being developed and improved. The observation method in combination with various types of polls is traditional for dialectology. With lingvopersonological research evolving, a method of inclusion is

implemented in field practice, i.e., inclusion into the language being of the speaker, which means establishing close relationships with the informant, regular long-term observation, and creation of a situation in the course of recording that is comfortable for the individual. Application of this method allows better approaching the situation of open recording in conditions where free speech is generated in the natural language environment and provides an opportunity to gather facts about both an individual's discourse and the speaker's personality (Ivantsova & Solomina 2014).

The formation of methods of inclusive observation may be promoted by kinship between collectors and dialect speakers. Thus, Timofeev had been recording the speech of his mother E. M. Timofeeva for more than 20 years (Timofeev 1971), and Lyutikova (1999) had been doing so with her mother's speech (V. M. Petukhova) for approximately the same length of time. However, in a number of dialectological projects absence of family relations was successfully compensated by confidential relations. Inclusion in the language being of the speaker was practiced by the Perm researchers who had been writing down A. G. Gorshkova's speech for about ten years (Skitova & Ogiyenko 1971) and by the Tomsk dialectologists who studied V. P. Vershinina's speech systematically for 24 years. Elements of this method were applied in other lingvopersonological projects, though the degree of confidentiality between informants and collectors was different in every case.

6. Aspects of research

Both traditional and relatively new areas of analysis have been applied to researching idiolects of individual dialect speakers.

Within dialectological traditions, phonetic and grammatical phenomena of the individual speech of peasants, typical to their dialects, are described: features of vocalism and consonantism, inflexions of content words, and some syntactic characteristics (originality of functions of auxiliary parts of speech, word compatibility, and specific types of sentences). The idiolect of the person who bears a folk speech culture is perceived in these cases as "a chip of the dialect". Conclusions are drawn

about its complete or incomplete coincidence with the dialect which is native for the individual (Timofeev 1971: 121-138; Lyutikova 1999: 20-40; Prokofieva 2012: 43-101). Phonetic and grammatical phenomena that characterize the author as a representative of a certain dialect group are also reconstructed in written texts of dialect speakers on the basis of analyzing deviations from literary norms. In addition, the way dialect speakers master written and literary language is considered through analysis of graphics, spelling, punctuation, and clichéd word combinations (Ossipov 1995; Batyreva 2011: 39-56).

The analysis of the lexical tier of the language system of dialect speakers focuses interest not only on locally limited lexical units (Timofeyev 1971, 2003; Lyutikova 1999, 2000; Nefyodova 1997, 2000, 2001), but also on the later tendency of the system analysis of the individual lexicon. The latter deals not only with lexemes that coincide with the literary language but also with those that don't coincide (Skitova & Ogienko 1971; Malysheva 2007; Ivantsova 2002, etc.). In works on lexicon, thematic classification of the individual lexicon is considered (Timofeyev 1971: 121-138). Many papers are devoted to separate groups of words in the idiolexicon: diminutives (Andreyeva & Gorlanova 1971), confessional nominations (Tolstova 2007), expressional and emotional elements (Nefyodova 1997), and some others. The description of the lexicon of a Siberian peasant V. P. Vershinina includes research of its all-Russian, colloquial, and dialect component, new and archaic vocabulary, nonce words and expressives, and also the main types of system relations of lexemes (motivational, alternative, synonymic and antonymic ones) (Ivantsova 2002: 36-160). Some authors also study quantitative characteristics of a dialect speaker's lexicon: its volume, the division of words into grammatical classes, a ratio of polysemantic and monosemantic units, and the rate of their use (Timofeyev 1971; Skitova & Ogiyenko 1971; Lyutikova 1999; Ivantsova 2002).

The fact that discourse research was developed at the beginning of the twenty-first century sparks keen interest in text created by a dialect speaker. General features of the structure of the text of the dialect language personality are considered (Ivantsova 2002: 180-250), as well as expressive means of the text – first of all, comparisons and metaphors (Lyutikova 1999; Ivantsova 2002; Volkova 2004), and a folklore component of the household discourse: proverbs, sayings, and humorous rhymes (*chastushkas*)

(Lyutikova 1999; Malysheva 2007: 115-130). The system of speech genres of an idiolect becomes a subject of close attention (Demeshkina 1997; Gyngazova 2001; Kazakova 2007). In written texts of dialect speakers – diaries, memoirs, and letters – their substantial and structural features are analyzed. They are compared with oral texts on a similar subject produced by the same informants. This allows revealing common features of these two types of texts – "colloquiality" of speech and contrasting features (Ossipov 1995; Russinova 2007).

A new aspect of studying of the dialect text is research of the reflection of personality over language as an important part of consciousness of the individual mirrored in the text. The metalanguage reflection in the speech of dialect speakers is thoroughly studied (Blinova 1984; Mikitina 1989; Rostova 2000) including that in separate idiolects (Sakharny & Orlova 1969; Lyutikova 1999; Ivantsova 2009). Forms of manifestation of metalanguage consciousness, the area of reflection, strategies of understanding of semantics of a dialectal word by the speaker, and particularities of assessing one's own speech and of the speech of people around are discovered.

Papers published in recent years have investigated the conceptosphere of specific representatives of folk oral culture based upon the data and the vocabulary test of dialectal language personalities. In a series of publications Gyngazova (2008, 2009, 2010, etc.) considered a system of key concepts of folk culture in the idiolect of the Siberian peasant V. P. Vershinina: HOUSE, LAND, LABOR, LIFE and DEATH, GOD, SIN, WAY, SPACE, BODY, SOUL, and others. Prokofieva (2012) analyzed such concepts as HOME, FAMILY, VILLAGE, WORLD, and GOD in the idiolect of the Altai dialect speaker A. V. Medvedeva. In the works of Ivantsova (2002, 2009, 2014), Volkova (2004), Russinova (2007) and Kuznetsova (2015) a particular picture of the world, the worldview, and the outlook of the dialect language personality were identified.

On the basis of the speech data of individual dialect speakers, the following problems were raised: problems of the creative beginning in the speech of ordinary native speakers (Lyutikova 1999; Prokofieva 2012), problems of detecting relic dialect features (Almukhamedova *et al.* 1986; Tolstova 2007), typological features of the dialect of the language personality (Ivantsova 2014), and the relation of language elements of an idiolect and a dialect (Gruzberg & Egoryeva 1969).

Dialect lexicography received a new impetus for its development. In the second half of the twentieth century not only the dialect speech of certain regions, but that of specific language personalities, became the object of lexicography. A special type of an idiolect dictionary of a dialect speaker with its different subtypes was created:

- a differential dictionary that includes only locally restricted units (Timofeev 1971; Lyutikova 2000);
- a non-differential dictionary, which reflects all-Russian dialect elements and dialect elements in a narrow sense on an equal footing (Tolstova 2004; Ivantsova 2006-2012);
- a general-type explanatory dictionary (includes all of the above);
- aspect dictionaries of an idiolect, the purpose of which is lexicographic representation of individual lexical classes of an idiolexicon, the means of expressive speech of an individual, and frequency characteristics of the text. Among the most recently published dictionaries are, *An Expressive Dictionary of a Dialect Personality* (Nefyodova 2001), *Phraseology of a Dialect Personality* (Timofeev 2003), and *An Idiolect Dictionary of Comparisons of a Longtime Siberian Resident* (Ivantsova 2005).

The range of aspects and problems of the study to be solved on the basis of data of individual dialect speakers is constantly expanding.

The study of individual speech of dialect speakers evolves from describing an idiolect as a "point representative of a dialect" to the dialectical understanding of the speakers' similarities and differences, and from analyzing only linguistic features of the individual to analyzing the personality of a native speaker, in which both a linguistic and an extralinguistic component are present. Researchers raise questions about conditions in which speech abilities of a speaker are formed in the dialect environment, about a speaker's mental attitudes and ethical and aesthetic preferences in a unique combination of typical features for traditional peasant communities and unique features of the individual.

It is frequent that studies are complex, multidimensional, and have a clearly marked lexicographic component.

7. Methods of research

Analysis of the practice of studying speech of specific representatives of folk speech culture has shown that scientists use well-known general scientific methods, purely linguistic ones as well as new task-oriented methods for the study of the phenomenon of the language personality. From the group of universal interdisciplinary methods the main method is that of scientific description, involving a systematic inventory of language units and their taxonomic characteristics for formal, substantial, and functional properties (Skitova & Ogienko 1971; Tolstova 2007; Batyreva 2011). The group of purely linguistic methods is dominated by the recently recognized independent lexicographic method, which is applied not only as a way to present the language material, but as an instrument of its analysis. The range of special lingvopersonological methods is represented by methods of speech portraiture and reconstruction of the language personality.

Speech portraiture rests on observable facts and is a story-like characteristic of the speech of the individual emphasizing its vivid features. In dialectology the most common type of a speech portrait includes brief biographical information about the informant, fragments of the informant's voice recordings, and the description of the non-literary features of idiolect phonetics, grammar, and vocabulary (Slesareva 1997; Oglezneva 2004; Kasatkin 2007; Baklanova 2008; Batyreva 2011).

The method of reconstructing the language personality is based on Karaulov's ideas and presupposes not only analyzing the linguistic means of the individual, but also reconstructing the worldview, objectives, interests, and the outlook of a person that for a direct observer are too difficult to access. This method focuses on the language personality of the past, embodied in a literary text. The object of the study may be the author of the text (writer) and also its characters (Karaulov 1987). The application of this method to the new object – a modern individual, including an ordinary native speaker who uses mostly spoken language – allows us to apply this method in dialectology. The result of such research is cognitive reconstruction of features of the worldview and

understanding of bearers of traditional folk speech culture in the works of Perm and Tomsk researchers.

The methodology of the largest projects to study the dialect language personality can be differentiated by a synthesis of the elements of the abovementioned methods.

8. Conclusion

Due to the development of an anthropocentric paradigm and the interaction of different linguistic disciplines, characteristic of modern science, a new object of study – a dialectal language personality – has appeared in dialectology. Its articulation took place for most of the last century and was characterized by:

- the appearance of works that not only describe a particular dialect as a whole, but also some of its speakers;
- a change of the focus of analysis from the speech of a particular dialect speaker as an illustration of usual characteristics of the dialect to research whose core of analysis is an individual who becomes a "starting point" in the study of a language;
- approbation of the method of inclusion in the linguistic being of the speaker in order to collect material under conditions as close as possible to the situation of spontaneous speaking of a dialect speaker;
- the creation of databases of scientific study of dialect language personalities based on a considerable number of records of oral speech and written texts of "naive authors";
- the creation of methods for studying the phenomenon of a dialect language personality.

Attention to the new object – the personality of the dialect speaker – is very significant for dialectology and lingvopersonology, at the junction of which studies of individual dialect speakers develop, as well as for linguistics in general.

In dialectology the study of the language personality allowed obtaining previously unknown information about quantitative and qualitative features of the vocabulary of peasants, particularities of speech culture in folk dialects, and originality of

communication in the dialect, and gave an impetus to developing new types of dialect dictionaries.

Analysis of individual speech of dialect speakers played a crucial role in shaping the general theory of lingvopersonology. It enriched the concept of a language personality, the spectrum of methods of collecting and analyzing data, the sources of speech data of ordinary speakers, and typological features of a bearer of folk speech culture. Theoretical generalizations from the study of individual representatives of dialects are now being implemented into researching other types of language personalities – speakers of the literary language, the urban colloquial language and jargon, elite personalities, and historical figures.

Studying the phenomenon of a dialect language personality becomes a "pilot site" for formulating and solving many general linguistic problems. In lingvopersonological works dialectologists have raised global questions about the genesis of the language personality (factors influencing its formation and development), and connections and differences in speech of the individual and the language of society to which he or she belongs, by defining the typical and the individual in an idiolect, features of national culture (the fundamental principle of which is a traditional national culture), and particularities of a discourse practice in the contemporary language community.

References

ALMUKHAMEDOVA, Z. *et al.* (1986) "Из наблюдений над речью Лыковых – старожилов на реке Абакан" [From observation of the Lykovs' speech – longtime residents on the Abakan River], *Phonetics and Writing*, Ustinov, 103-115.

ANDREYEVA, L. & N. GORLANOVA (1971) "Существительные с уменьшительно-ласкательными суффиксами в народно-разговорной речи" [Nouns with diminutive-hypocoristic suffixes in folk conversational speech], *Living word in the Russian speech of Prikamye*, Perm, Issue 2, 71-78.

ASAHI, Yoshiyuki (2009) "Cookbook method and koine-formation: A case of the Karafuto dialect in Sakhalin", *Dialectologia*, 2, 1-21.
<http://www.publicacions.ub.edu/revistes/dialectologia2>

BAKHTIN, M. (1996) *Проблема речевых жанров* [The problem of speech genres], in M. Bakhtin, *Collected works*, Moscow, Russian dictionaries, 159-206.

BAKLANOVA, I. (2008) "Речевой портрет диалектоносителя (на материале пермских говоров)" [Speech portrait of a dialect speaker (on the material of Perm dialects)], *Russian in Action*. Glazov, 123-127.

BATYREVA, L. (2011) *Речевой портрет провинциала (на материале частной переписки 1940-1945 гг. палехского крестьянина А.Н. Савельева)* [Speech portrait of a provincial (on the material of private correspondence of 1940-1945 of A. N. Savelyev, a Palekh peasant)], Shuya, 175.

BAUDOUIN DE COURTENAY, J. (1963) *Избранные труды по общему языкознанию,* Т. 1–2. [Selected works on general linguistics (in 2 vol.)], Moscow: Publishing House of USSR Academy of Sciences.

BLINOVA, O. (1984) *"*Носители диалекта – о своем диалекте (об одном из источников лексикологического исследования)*"* [Dialect speakers about their dialect (about one of the sources of lexicological studies)]*, Siberian dialects of Russian*, Tomsk**,** 3-15.

BOGIN, G. (1984) *Модель языковой личности в ее отношении к разновидностям текстов* [Model of language personality in its relation to varieties of texts], abstract of doctoral dissertation, Leningrad, 31 p.

BUSCHMANN, E. & W. VON HUMBOLDT (2000) *Wörterbuch der Mexicanischen Sprache*, Manfred Ringmacher (ed.), Paderborn: Schöningh.

DEMESHKINA, T. (1997) "Жанровые формы диалектной речи (на материале языка личности)" [Genre forms of dialect speech (on the material of the language of a personality)], *Phenomenon of variability*. Kemerovo, 122-129.

FELDE, O. (2010) "«Была ужасная пора, о ней свежо воспоминанье…» Устные рассказы жительницы села Степной Баджей Манского района Красноярского края Семенковой Галины Васильевны" ["There was an awful time, fresh is the reminiscence about it …" Oral stories of Semenkova Galina Vasilyevna, inhabitant of the Stepnoy Badzhey village of the Mansky Region of the Krasnoyarsk Krai], *Siberia Lingua. Scientific journal.* Krasnoyarsk, No 2, 184-219.
<http://ifiyak.sfu-kras.ru/sites/language.institute.sfu-kras.ru/files/Siberia_Lingua_no_2.pdf>

FILLMORE, Charles J., Daniel KEMPLER & William S-Y. WANG (eds.) (1979) *Individual differences in language ability and language behavior*, New York: Academic Press, 346.

GOLEV, N. (2004) Лингвоперсонологическая вариативность языка [Individual variability of language], *Bulletin of Altai State University,* 4, 41-45.

GRUZBERG, L. & M. EGORYEVA (1969) "О соотношении языка говора и речи одного носителя говора (на материале вводных элементов)" [On the ratio of a dialect and the speech of one dialect speaker (on the material of introductory elements)], *Living word in the Russian speech of Prikamye*, Perm, 65-75.

GYNGAZOVA, L. (2001) "О речевом жанре воспоминания (на материале языка личности)" [On a speech genre of reminiscence (on the material of the language of a personality)], *Modern directions of functional linguistics*, Tomsk, 167-174.

GYNGAZOVA, L. (2008) "Физическое и духовное пространство в дискурсе носителя традиционной культуры" [Physical and spiritual space in the discourse of a bearer of traditional culture], *Pictures of the Russian world: Spatial models in the language and in the text*, Tomsk: UFO-PLUS, 78-109.

GYNGAZOVA, L. (2009) "Интерпретационный потенциал соматизмов в описании картины мира языковой личности диалектоносителя" [Interpretative potential of somatisms in the description of the picture of the world of a dialect speaker's language personality], *Vestnik of Tomsk State University, Philology*, Tomsk. No 1(5), 13-22.

GYNGAZOVA, L. (2010) "Метафорическое миромоделирование в дискурсе языковой личности" [Metaphorical world modeling in the discourse of the language personality], *Vestnik of Tomsk State University, Philology*, Tomsk, 1(9), 7−11.

IVANTSOVA, E. (2002) *Феномен диалектной языковой личности* [Phenomenon of the dialect language personality], Tomsk: Tomsk State University Publishing House, 312.

IVANTSOVA, E. (2005) *Идиолектный словарь сравнений сибирского старожила* [The Idiolect Dictionary of Comparisons of a Siberian Longtime Resident], Tomsk: Tomsk State University Publishing House, 162.

IVANTSOVA, E. (ed.) (2006-2012) *Полный словарь диалектной языковой личности* [Unabridged Dictionary of a Dialect Language Personality], Vol. 1-4, Tomsk: Tomsk State University Publishing House.

IVANTSOVA, E. (2009) *Метаязыковое сознание диалектной языковой личности* [Metalanguage consciousness of the dialect language personality], Everyday language consciousness: Ontological and gnoseological aspects, in N. D. Golev (ed.), Part. 1. Kemerovo − Barnaul: Altai State University Publishing House, 321-356.

IVANTSOVA, E. (2010) *Лингвоперсонология: Основы теории языковой личности* [Lingvopersonology: Foundations of the theory of the language personality], Tomsk: Tomsk State University Publishing House, 160.

IVANTSOVA, E. (2014) *Исследование типологических черт диалектной языковой личности* [Research of typological features of the dialect language personality] Language personality: Modeling, typology, portraiture (Siberian lingvopersonology), in N. D. Golev & N. N. Shpilnaya (eds.), Moscow: Lenand, 309-348.

IVANTSOVA, E. & E. SOLOMINA (2014) "Об эффективных методах записи спонтанной устной речи при изучении языковой личности" [On effective methods of record of spontaneous oral speech when studying the language personality], *Vestnik of Tomsk State University. Philology*, No 3(29), 14-27.

JOHNSTONE, Barbara (1996) *The linguistic individual. Self-expression in language and linguistics,* New York/Oxford: Oxford University Press, 210.

KARAULOV, Yu. (1987) *Русский язык и языковая личность* [The Russian language and the language personality], Moscow: *Science*, 262.

KASATKIN, L. (2007) "Фрагмент языкового портрета донской казачки" [A fragment of the language portrait of a Don Cossack woman], *Language in action*, Moscow, 231-240.

KAZAKOVA, O. (2007) *Диалектная языковая личность в жанровом аспекте* [The dialect language personality in the genre aspect], Tomsk: Tomsk Polytechnic University Publishing House.

KUZNETSOVA, S. (2015) *Вербализация чувственного восприятия как отражение перцептивной картины мира диалектной языковой личности* [Verbalization of perception as a reflection of the perceptive world view of the dialect speaker], PhD dissertation, Tomsk.

LYUTIKOVA, V. (1999) *Языковая личность и идиолект* [The language personality and an idiolect], Tyumen: Tyumen University Publishing House, 188.

LYUTIKOVA, V. (2000) *Словарь диалектной личности* [The Dialect Personality Dictionary], Tyumen: Tyumen University Publishing House, 188.

MALYSHEVA, V. (2007) *Идиолексикон сквозь призму лексикографии* [Idiolexikon through a lexicography prism], Issue 1, Perm, 136.

MARKELOV, V. (2000) "Письма Агафьи Лыковой" [Agafya Lykova's letters], *Russian Speech*, No 4, 65-73.

NEFYODOVA, E. (1997) "Экспрессивный потенциал языковой (диалектной) личности" [Expressional potential of the language (dialect) personality], *Issues of Russian Linguistics*, *Issue 7*, *Russian dialects: History and modern times*, Moscow, 220-229.

NEFYODOVA, E. (2000) "Идиолект как источник диалектного варьирования" [An idiolect as a source of a dialect variation], *Materials and research on Russian dialectology,* Moscow, Issue 1 (7), 251-262.

NEFYODOVA, E. (2001) *Экспрессивный словарь диалектной языковой личности* [Expressional Dictionary of the Dialect Language Personality], Moscow: Moscow State University Publishing House, 144.

NEROZNAK, V. (2003) "Лингвокультурология и лингвоперсонология" [Cultural linguistics and Lingvopersonology], *The Tower of Babel-2: Word. Text. Culture. Reading 2002-2003*, Moscow, 13-17.

NIKITINA, S. (1989) *"*Языковое сознание и самосознание личности в народной культуре*"* [Language Consciousness and Self-Consciousness in Folk Cuture]. *Language and Personality*, Moscow, 34-40.

OGLEZNEVA, E. (2004) *"Социально-речевой портрет современного диалектоносителя (на материале речи М.В. Хлыстова, жителя с. Чернавка Свободненского района Амурской области)"* [The social and speech portrait of a modern dialect speaker (on the material of the speech of M. V. Hlystov, the inhabitant of the Chernavka village of the Svobodnensky Region of the Amur Region)], *Folk speech of Primorye*, Blagoveshchensk, 85-95.

OSSIPOV, B. (ed.) (1995) *Автобиографические записки сибирского крестьянина В.А. Плотникова: Публикация и исследование текста* [Autobiographical notes of a Siberian peasant V. A. Plotnikov: Publishing and research of the text] Preparation, foreword and commentary, Omsk: Omsk State University Publishing House.

OSTHOFF, H. & K. BRUGMAN (1881) *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*, Erster Theil, Leipzig <https://archive.org/details/morphologischeu00bruggoog>.

PAUL, H. (2002) *Prinzipien der Sprachgeschichte*, 10th Unrev. edition, Tübingen: Niemeyer.

PETKOV, S. (1983) *Език и личност* [Language and personality], Influence, Blagoevgrad.

POUND, Louis (1947) "The American Dialect of Charles Dickens", *American Speech*, vol. 22, No. 2, 124-130.

PROKOFIEVA, E. (2012) *Диалектная языковая личность на Алтае* [A dialect language personality in Altai], PhD dissertation, Barnaul, 235.

ROSTOVA, A. (2000) *Метатекст как форма экспликации метаязыкового сознания* [Metatext as a form of explication of metalanguage consciousness], Tomsk: Tomsk State

University Publishing House.

RUSSINOVA, I. (ed.) (2007) Дневник М.П. Сусловой (1981-1985 гг.): Публикация и исследование текста [M. P. Suslova's diary (1981-1985): Publication and research of the text], Perm, 264.

SAKHARNY, L. & O. ORLOVA (1969) "Типы употребления в речи нескольких вариантов одной гиперлексемы (опыт психолингвистического анализа текста)" [Types of usage of several options of one hyperlexeme in speech (experience of the psycholinguistic analysis of the text)], *Living word in the Russian speech of Prikamye*, Perm. Issue 1, 83-113.

SAPIR, Edward (1949) *Selected writings of Edward Sapir in language, culture, and personality,* David G. Mandelbaum (ed.), Berkeley: University of California Press.

SHAKHMATOV, A. (1914) *Описание лекинского говора Егорьевского уезда Рязанской губернии* [Description of the Lekinsky dialect of the Egoryevsky District of the Ryazan Province], St. Petersburg, 48.

SKITOVA, F. & E. OGIYENKO (1971) "Из наблюдений над словарным запасом одного человека" [From observation of the lexicon of one person], *Living word in the Russian speech of Prikamye*, Perm. Issue 2, 26-39.

SLESAREVA, G. (1997) "Звуковой портрет Агафьи Лыковой" [A sound portrait of Agafya Lykova], *Proceedings of the International Council of Experts in Russian Language and Literature in Krasnoyarsk*, vol. 1, 173-174.

TIMOFEYEV, V. (1971) *Диалектный словарь личности* [A dialect dictionary of a personality], Shadrinsk, 141.

TIMOFEYEV, V. (2003) *Фразеология диалектной личности: Словарь* [Phraseology of a dialect personality: Dictionary], Shadrinsk: Shadrinsk Pedagogical Institute Publishing House, 176.

TOLSTOVA, G. (2004) *Словарь языка Агафьи Лыковой* [A Dictionary of Agafya Lykova's language], Krasnoyarsk.

TOLSTOVA, G. (2007) *Старообрядческая конфессиональная лексика в письменной речи Агафьи Лыковой* [Old Believers' confessional lexicon in Agafya Lykova's written language], abstract of PhD dissertation, Kemerovo, 28.

VINOGRADOV, V. (1980) *Избранные труды. О языке художественной прозы*. [Selected works. About language of fiction], Moscow, Science.

VOLKOVA, N. & E. SAFONOVA (2010) "Языковая личность в диалектном пространстве" [A language personality in the dialect environment], *Vilniaus pedagoginis universitetas,* vol. 12, 55-59.

VOLKOVA, T. (2004) *Сравнения в речи диалектной языковой личности* [Comparisons in the speech of the dialect language personality], PhD dissertation, Tomsk, 237.

WEISGERBER, L. (1929) *Muttersprache und Geistesbildung*, Göttingen: Vandenhoeck Ruprecht.

Dialectologia 17 *(2016), 109-122.*
ISSN: 2013-2247

Received 30 November 2014.
Accepted 12 March 2015.

REDUPLICATION WITH FIXED SEGMENTISM IN CENTRAL SARAWANI BALOCHI

Bahareh SOOHANI[1], Abbas Ali AHANGAR[2] & Marc van OOSTENDORP[3]

Leiden University, the Netherlands[1]*

University of Sistan and Baluchestan, Iran[2]*

Meertens Institute, the Netherlands[3]*

soohani@hum.leidenuniv.nl/ aaliahangar@ English.usb.ac.ir/

Marc.van.Oostendorp@meertens.knaw.nl

Abstract

The Central Sarawani dialect of Balochi (Indo-European, Iran), has a number of reduplicative patterns. One of these is an augmentative pattern that we will refer to as '*m*/*p*-reduplication' and which instantiates an example of 'morphological fixed segmentism' in the sense of Alderete et al. (1999). The present study examines this type of reduplication in Sarawani Balochi based on Optimality Theory (OT). The linguistic corpus relies on an original fieldwork through the purposeful recording of speech gathered through interview with 10 male and female language consultants with different social backgrounds. The research findings show that this type of augmentative reduplication can be represented by ranking the following constraints: OCP, FAITH-AFFIX, MAX-BR, *ONS/N, IDENT-BR (lab), and VOP. More interestingly, however, this segment is not completely fixed: in most cases it is *m*, but this is not true when the stem itself contains *m*, it is *p* instead.

Keywords

augmentative reduplication, fixed segmentism, optimality theory, constraints, onset sonority hierarchy

* Leiden University, PO Box 9500, 2300 RA Leiden.
* English language and literature department; Faculty of literature and humanities; Sistan and Baluchestan University, Zahedan, Iran.
* Meertens Institute, PO Box 94264, 1090 GG Amsterdam.

REDUPLICACIÓN CON SEGMENTISMO FIJO

EN SARAWANI BALOCHI CENTRAL

Resumen

El dialecto Sarawani Balochi central (Indoeuropeo, Irán), tiene una serie de patrones reduplicativos. Uno de ellos es un patrón aumentativo al cual nos referiremos como 'm/p-reduplicación', y que crea un ejemplo de 'segmentismo morfológico fijo' en el sentido de Alderete *et al.* (1999). El presente estudio examina este tipo de reduplicación en Sarawani Balochi basándose en la Teoría de Optimidad (TO). El corpus lingüístico se basa en un trabajo de campo original a través de la grabación expresa del habla obtenida a través de entrevistas a 10 informadores hombres y mujeres de distintos orígenes sociales. Los resultados de la investigación muestran que este tipo de reduplicación aumentativa puede representarse mediante la clasificación de las siguientes restricciones: OCP, FAITH-AFFIX, MAX-BR, *ONS/N, IDENT-BR (lab), y VOP. Curiosamente, sin embargo, este segmento no es completamente fijo: en la mayoría de los casos es *m*, pero esto no es cierto cuando la raíz en sí contiene *m*, entonces es *p*.

Palabras clave

reduplicación aumentativa, segmentismo fijo, teoría de optimidad, restricciones, jerarquía de la sonicidad inicial

1. Introduction

Augmentative reduplication with fixed segmentism requires copying of the base elements coupled with introducing a fixed segment. The added segment is an affix that is realized simultaneously with the reduplicative copy, and overwrites a portion of the reduplicant (McCarthy & Prince 1986, 1990).

Augmentative reduplication with fixed segmentism in Central Sarawani Balochi (henceforth CSB), as in other languages with this phenomenon, among various reduplicative patterns (cf. Moradi 2012), is an example of rhyming patterns. We refer to this kind of reduplication in CSB as *m*/*p*- reduplication, which illustrates the most productive type of reduplication in this dialect:

Dialectologia 17 *(2016), 109-122.*
ISSN: 2013-2247

(1) *m*/*p*- Reduplication

Base		Reduplicative form	
a. *bætʃæk*	'boy'	*bætʃækmætʃæk*	'boy and so forth'
b. *gʊk*	'cow'	*gʊkmʊk*	'cow and the like'
c. *mu*ː*d*	'table'	*mu*ː*dpu*ː*d*	'hair and the like'

In the present article, our analysis will be based on Optimality Theory, a theory of constraint interactions in grammar (Prince & Smolensky 1993, McCarthy & Prince 1993a, b). In our OT analysis, we will show that '*m*/*p*-augmentative reduplication' is an example of melodic over-writing (McCarthy & Prince 1986, 1990) and Alderete *et al.* (1999) as 'Morphological type of fixed segmentism'.

This article proceeds as follows: §2 introduces the language background; §3 deals with the theoretical framework employed; §4 provides a description and an analysis of the linguistic data; and finally §5 represents the conclusion.

2. Language background

Balochi is spoken in south-western Pakistan, and by a large number of people in Karachi. It is also spoken in south-eastern Iran, in the province of Sistan and Baluchestan, and by Baloch who have settled in the north-eastern province of Khorasan and Golestan. It is, furthermore, spoken by small communities in Afghanistan, in the Gulf States, in the Marw/Marie region of Turkmenistan, in India, East Africa and today also by a considerable number of Baloch in North America, Europe and Australia (Jahani & Korn 2009). The number of Balochi speakers is estimated between 5- 8 million (Jahani 2001: 59)

Jahani & Korn (2009: 636) divide the main dialects of Balochi into Western, Southern, and Eastern. They declare this is a very broad dialect division, within which further dialect demarcations can be made. Some dialects do not easily fit any of these groups. This is true, for example, of the dialect spoken in Iranian Sarawan, which shows transitional features between Western and Southern.

The dialect of Sarawani Balochi is spoken in the valley from Zangiyan, Dezzak, the central town of Sarawan (including Shastun, Sarjo and Bakhshan) up to Hoshshak, Gosht, Jalk, Kallagan, Naug as well as in Paskoh and Seb in the Soran valley and past the Siyahan mountain range but not in the rest of Soran approximately the same dialect is spoken. This dialect is called Central Sarawani by Baranzehi (2003). Dehwar with its surrounding areas and almost the whole Soran Valley except for Paskoh and Seb can be classified as another dialect; this dialect is called Sorani/Dehwari (Baranzehi 2003).

3. Theoretical considerations

3.1 Basic concepts of Optimality Theory

The central idea of OT is that surface forms of language reflect resolutions of conflicts between competing demands or constraints. A surface form is 'optimal' in the sense that it incurs the least serious violations of a set of violable constraints, ranked in a language-specific hierarchy. Constraints are universal and violable, and directly encode markedness statements and principles enforcing the presentation of constraints. A language differs in the ranking of constraints, giving priorities of some constraints over others. In fact, the optimal output form arises from competition of markedness and faithfulness constraints. Faithfulness constraints require that output be the same as their lexical input, in other words, faithfulness constraints oppose changes, while markedness constraints trigger changes (Prince & Smolensky 1993, McCarthy & Prince 1993a, b). In addition, 'faithfulness constraints state their requirements about input-output relations in term of correspondence' (Kager 1999: 194).

Reduplication is a phenomenon which involves phonological identity between the 'reduplicant' and the 'base' (Kager 1999: 194). McCarthy & Prince (1994) give the definition of 'base' and 'reduplicant' as paraphrased in Kager (1999: 202):

(2) 'The 'reduplicant' is the string of segments that is the phonological realization of some reduplicative morpheme RED, which is phonologically empty.
The 'base' is the output string of segments to which the reduplicant is attached, more specifically:

- For reduplicative prefixes, it is the following string of segments.
- For reduplicative suffixes, the preceding string of segments.'

3.2 Fixed Segmentism

Reduplication refers to a word formation process that can result in an identical copy of the base, or not (Urbanczyk 2007: 474). In addition to being composed of segments from the base, reduplication can also contain fixed segments. Following the work of McCarthy & Prince (1986, 1990), Alderete *et al.* (1999) argue that there are two distinctive types of reduplication with fixed segmentism: default segmentism and melodic overwriting. In the former a default segment is phonologically motivated and it is generally the least marked and also frequently the epenthetic segment of a language.

On the other hand, following McCarthy & Prince (1986, 1990), Yip (1992) and Alderete *et al.* (1999) have discussed that the overwriting string is an affixial morpheme which is relatively marked segments that replace segments from the base, as with the *schm*-reduplication pattern in English: *table-schmable*. Moreover, Alderete *et al*. (1999: 357) illustrate the properties of morphological fixed segmentism based on affixation as follows:

a) Faithfulness: fixed segments may form marked structure and be in contrast with other fixed segments.

b) Alignment: fixed segments may be left-aligned, right-aligned or infixed.

c) Context-sensitivity: fixed segments may alternate by suppletion or allomorphy.

In what follows we shall analyze the *m/p*-reduplication in CSB with a constrain-based approach as well as the theory suggested by Alderete *et al.* (1999).

4. Analysis of *m*/*p* – Reduplication in CSB

Display (4) provides examples of the augmentative reduplication forms in CSB, which highlights frequency, size or intensity. In CSB, the overwriting morpheme is generally *m*-this overwriting morpheme can, however, alternate by suppletion or allomorphy just like other affixes. So CSB selects the alternant *p*- when the word already starts with *m*- , like the forms in (4b):

(4) *m* /*p*- augmentative Reduplication

	Base		Reduplicative form	
a.	*ʧokk*	'child'	*ʧokkmokk*	'child and so forth'
	kotʃæk	'dog'	*kotʃækmotʃæk*	'dog and so forth'
	gʊk	'cow'	*gʊkmʊk*	'cow and so forth'
	kæt	'room'	*kætmæt*	'book and the like'
	lehip	'blanket'	*lehipmehip*	'blanket and the like'
	potʃtʃ	'cloth'	*potʃtʃmotʃtʃ*	'cloth and the like'
b.	*moʃk*	'mouse'	*moʃkpoʃk*	'mouse and so forth'
	mʊr	'ant'	*mʊrpʊr*	'ant and the like'
	mu:*d*	'hair	*mu*:*dpu*:*d*	'hair and so forth'

As data in (4a-b) show the overwriting morpheme contain *m*- or *p*- as a prefix. Moreover; the reduplicant is a suffix. Within OT, Generalized Alignment (McCarthy & Prince 1993a, cited in Ussishkin 2007: 458) provides a framework for analyzing morpheme position. So, the following two constraints inflict alignment restrictions on the affixal morpheme and the reduplicant respectively:

(5) ALIGN-L (*m*/*p*-, RED)

The left edge of *m*/*p*- is aligned to the left edge of a reduplicant.

(6) ALIGN-R (RED, BASE)

The right edge of reduplicant is aligned to the right edge of a base.

Dialectologia 17 *(2016), 109-122.*
ISSN: 2013-2247

Based on our explanations given so far, it is clear that the prefix m/p- precedence over the reduplicant, in other words, it affects the reduplicant and not the base. Therefore, the presence of an overwriting morpheme indicates that faithfulness to overwriting morpheme has taken precedence, through ranking, over base-reduplicant (BR) faithfulness constraint. A constraint forcing the realization of affix material known as FAITH-AFFIX (Ussishkin 2007: 467).

(7) FAITH-AFFIX

Every morpheme in the input has to show up in the output.

(8) MAX-BR

Every element of Base has a correspondent in Reduplicant.

('No partial reduplication')

(9) MAX-IO

Input segment must have output correspondence

(No deletion)

Tableau (1) shows the effect of high-ranking FAITH-AFFIX in forming the *m*-reduplicant form from the base like in *gʊkmʊk* 'cow and the like'.

/gʊk-RED-m/	FAITH-AFFIX	MAX-IO	MAX-BR	ALIGN-L(*m*-,RED)	ALIGN-R(RED, BASE)
a. gʊk-gʊk	* !	*		*	
☞ b. gʊk-mʊk			* !		
c. mʊk-gʊk		*!	*		*
d. mʊk-mʊk		*!		*	

Tableau (1)

Forming *m*-reduplicant from the base

FAITH-AFFIX>> MAX-IO>> MAX-BR>>ALIGN-L (*m*-, RED), ALIGN-R (RED, BASE)

The optimal candidate [gʊk-mʊk], incurs a violation of MAX-BR while it satisfies other constraints. A candidate such as *mʊk-gʊk*, which faithfully realizes the input affix *m*-, is eliminated due to its violation of MAX-IO and MAX-BR, which are against deletion

and this candidate does not achieve perfect alignment of the base to the right edge as well.

In the case of *p*-reduplication in examples (4b), *p*- is an alternation of overwriting affix when the word already starts with *m*-. Therefore; the output like *mu*:*d-mmu*:*d* is ungrammatical, since it violates the Obligatory Contour Principle (OCP). But, why is *p*- an alternate affix and not other segments like *b*- or even *t*- (since coronals are universally less marked)? To find an answer for this kind of question and to make an analysis for *p*-reduplication in CSB based on OT, we should introduce the concept of Sonority Sequencing Generalization (SSG) based on Zec (2007).

Zec (2007: 187) states the Sonority Sequencing Generalization based on Selkirk (1984: 116) as the following statement:

(10) *'Sonority Sequencing Generalization (SSG*)
For every pair of segments *s* and *z* in a syllable, *s* is less sonorous than *z* if
a) (i) *s* < *z*< Nucleus
or (ii) Nucleus > *z* > *s*
or b) (i) *s* < *z* and *z* is the nucleus
or (ii) *z* > *s* and *z* is the nucleus'

Moreover; some restrictions may impose on the rise or fall in sonority that go beyond the minimal requirements of SSG by constraints on sonority distance, Prince & Smolensly's (2004) natural hierarchy of margins is based on these constraints on sonority distance. The best margins are obstruent followed by nasal and liquids. The hierarchy of onsets based on Prince & Smolensky (1993) is as follows (as cited in Zec, 2007: 188):

(11) *ONS/L >> *ONS/N >> *ONS/O

This hierarchy illustrates that the preference for onset is the lowest sonorous segments, so the least marked onsets are obstruent, and the most marked onsets are liquids.

In addition, the unmarked value for the feature [voice] in obstruent is [-voice], as stated in Voice Obstruent Prohibition (Kager 1999: 40), which is accompanied with the other two constraints relevant to the *p*-reduplication pattern.

Now, it will be clear that why the optimal reduplicant candidate for the input like *mu*:*d* in CSB is *mu*:*dpu*:*d*. First, as Prince & Smolensky's (2004) hierarchy of onset yields, an obstruent is the least marked segment, and [-voice] is the unmarked value in obstruent. Second, the presence of an alternation overwriting affix *p*- indicates that labiality faithfulness constraint needs to be in our constraint ranking. In sum, the following relevant constraints should be considered for *p*-reduplication based on OT:

(12) OCP

At the melodic level, adjacent identical elements are prohibited.

(13) *ONS/N

Word-initial syllables may not begin with nasal.

(14) VOP

*[+ voice, -son]

No obstruent must be voiced.

(15) IDENT-BR (lab)

Correspondent onsets are identical in their specification for bilabiality.

The FAITH-AFFIX and MAX-BR as stated earlier are accompanied with the other four constraints introduced above, are ranked in the following way in forming *p*-reduplicant from the base:

/mu:d-RED-m/	OCP	FAITH-AFFIX	MAX-IO	*ONS/N	MAX-BR	VOP	IDENT-BR(lab)
a. mu:d-mu:d			*	**!			
☞ b. mu:d-pu:d			*	*	*!		
c. mu:d-mmu:d	*!			***			
d. mu:d-bu:d			*	*	*	*!	
e. mu:d-tu:d			*	*	*		*

Tableau (2). OCP>>FAITH-AFFIX>>MAX-IO>>*ONS/N>>MAX-BR>> VOP, IDENT-BR (lab)

In this tableau, the optimal candidate from input /muːd-RED-m/ is [muːd-puːd], with same place of articulation to affix *m-.* Although it violates *ONS/N (since the base onset is a bilabial nasal [m]), this candidate is optimal because it avoids the violation of IDENT-BR(lab) , as shown by the comparison with suboptimal candidate [muːd-tuːd] in (2e), and also it avoids violation of VOP, as shown by the comparison with the suboptimal candidate [muːd-buːd] in (2d).

5. Conclusion

As discussed, we examined *m*/*p*-reduplication in Central Sarawani Balochi as a type of reduplication which is used to signal augmentative meaning. We proposed, based on our Optimality theoretical analysis, that m/p-reduplication is an example of Alderete *et al.* (1999) morphological fixed segemntism. Then we argued that prefix *m*- is considered as overwriting morpheme. Therefore, the presence of such a morpheme indicated that IO faithfulness to the overwriting morpheme has taken precedence, through ranking over the MAX-IO faithfulness constraint. Tableau1 established this result formally. Moreover; we regarded morpheme *m*- as a prefix which selected the alternate *p*- when the word has already started with *m*. Based on SGG and constraints on onset sonority hierarchy, we explained that an alternation affix for *m*- should be a voiceless obstruent, but bilabial. Tableau 2 demonstrated the ranking for relevant constraints to form an optimal output with *p*- as a fixed segment.

References

ALDERETE, John, Jill BECKMAN, Laura BENUA, Amalia GNANADESIKAN, John MCCARTHY & Suzanne URBANCZYK (1999) "Reduplication with fixed segmentism", *Linguistic Inquiry*, 30, 327-364.

BARANZEHI, Adam Nader (2003) "The Sarawani dialect of Balochi and Persian influence on it", in Carina Jahani & Agnes Korn (eds), *The Baloch and Their Neighbors*, *Ethnic and Linguistic*

Contact in Balochestan in Historical and Modern Times, Wiesbaden: Reichert Verlag, 75-111.

JAHANI, Carina (2001) "Balochi", in J. Garry & C. Rubino (eds.), *Facts about the World's Languages: An Encyclopedia of Word's Major Languages, Past and Present,* New York/ Dublin: H. W. Wilson, 59-64.

JAHANI, Carina & Agnes KORN (2009) "Balochi", in Gernot Windfuhr (ed.), *The Iranian languages*, London/New York: Routledge, 634-692.

KAGER, Rene (1999) *Optimality Theory*, UK: Blackwell.

MCCARTHY, John & Alan PRINCE (1986) "Prosodic Morphology", Ms., University of Massachusetts, Amherst, and Brandeis University, Waltham, Mass. [Issued in 1996 as Technical Report 32, Rutgers University Center for Cognitive Science.]

MCCARTHY, John & Alan PRINCE (1990) "Foot and word in Prosodic Morphology: The Arabic broken plural", *Natural Language & Linguistic Theory*, 8, 209-283.

MCCARTHY, John & Alan PRINCE (1993a) "Generalized Alignment", in G. E. Booij and J. van Marle (eds.), *Yearbook of Morphology*, *1993*, Dordrecht: Kluwer, 79-153.

MCCARTHY, John & Alan PRINCE (1993b) *Prosodic Morphology I: constraint interaction and satisfaction*, University of Massachusetts, Amherst and Rutgers University. [To appear as Technical report no. 3, Rutgers Universiry Center for Cognitive Science. Cambridge, Mass.: MIT Press.]

MCCARTHY, John & Alan PRINCE (1994) "Two lectures on prosodic morphology", *Handouts of two lectures at OTS/HIL workshop on Prosodic Morphology*, Utrecht University, July 1994. [ROA-59, http://ruccs.rutgers.edu/roa.html].

MORADI, Mohammadbakhsh (2012) *Morphological Description of Sarawani Balochi Dialect of Seb*, Unpublished MA Thesis, Azad University of Fars Research Center.

PRINCE, Alan & Paul SMOLENSKY (1993) Optimality Theory: Constraint interaction in generative grammar. Unpublished ms., Rutgers University & University of Colorado, Boulder. Published 2004, Malden, MA & Oxford: Blackwell.

SELKIRK, Elizabeth O. (1984) "On the major class features and syllable theory", in Mark Aronoff & Richard T. Oerhle (eds), *Language sound structure: studies in phonology and presented to Morris Halle by his teacher students,* Cambridge, MA: MIT Press.

URBANCZYK, Suzanne (2007) "Reduplication", in Paul de Lacy (ed.), *The Cambridge Handbook of Phonology*, Cambridge: Cambridge University Press, 473-493.

USSISHKIN, Adam (2007) "Morpheme position", in Paul de Lacy (ed.), *The Cambridge Handbook of Phonology*, Cambridge: Cambridge University Press, 457-472.

YIP, Moria (1992) "Feature geometry and cooccurrence restrictions", *Phonology*, 6, 349-374.

ZEC, Draga (2007) "The syllable", in Paul de Lacy (ed.), *The Cambridge Handbook of Phonology*, Cambridge: Cambridge University Press, 160-194.

Dialectologia 17 *(2016), 123-150.*
ISSN: 2013-2247

Received 15 November 2014.
Accepted 15 March 2015.

RUSSIAN IN BESERMAN ORAL DISCOURSE: CODE-MIXING AND BORROWING[1]

Maria USACHEVA & Olga BIRYUK

Moscow State University*

mashastroeva@gmail.com/ biryuk.olga@gmail.com

Abstract

The paper is devoted to the influence which Russian as a dominant language exerts on Beserman dialect of Udmurt. Analysis is based on a corpus of dialogues recorded and transcribed in Shamardan (a Beserman village in Udmurtia, Russian Federation). It is shown that Russian influence in forms of code-mixing and borrowing can be seen at all language levels: phonetics, inflectional morphology, vocabulary, syntax, discourse. It is stated that the changes could result in a mixed idiom, but the dialect is likely to die because it is not passed to next generations anymore.

Keywords

code-mixing, borrowing, Beserman, Russian, Udmurt

EL RUSO EN EL DISCURSO ORAL DEL BESERMAN: CAMBIO DE CÓDIGO Y PRÉSTAMOS

Resumen

El artículo estudia la influencia que el ruso como lengua dominante ejerce sobre el Beserman, dialecto del udmurto. El análisis se basa en un corpus de diálogos grabados y transcritos en Shamardan (un pueblo

[1] The data the current work is based on were gathered during fieldwork financed by RFH, grant № 16-24-17003 "Integral analysis of noun phrases in Finno-Ugric languages: support of reference and encoding information structure of the utterance". The analytical work was supported by RFBR, grant № 16-06-00536 "Syntax-semantic interface in Uralic and Altaic languages".

* 119991, Lomonosov Moscow State University, Leninskiye gory, GSP-1, Moscow, Russian Federation.

dónde se habla Beserman, en Udmurtia, Federación Rusa). Se muestra que la influencia rusa en forma de cambio de código y de préstamo se puede observar en todos los niveles de la lengua: la fonética, la morfología flexiva, el vocabulario, la sintaxis y el discurso. Se afirma que los cambios experimentados podrían dar lugar a una lengua mixta, aunque el dialecto es probable que muera ya que no se transmite de generación en generación.

Palabras clave

cambio de código, préstamos, Beserman, ruso, udmurto

1. Introduction

This article concerns the Beserman dialect of Udmurt. Besermans is a relatively small (according to the All-Russian population census, there are 2201 people identifying themselves as Beserman) ethnic group occupying the basin of Cheptsa river and the Kirov region of the Russian Federation. Besermans consider themselves to be a nationality different from Udmurts and to speak a unique language. But in the scientific literature the Beserman idiom is considered to be a dialect of Udmurt language which is characterized by an unusual combination of specifically Beserman language phenomena (concentrated in vocabulary and phonetics) with certain traits of northern and southern Udmurt dialects (mostly morphological and phonetical) (see Teplyashina 1970, Kelmakov 1998, Ljukina 2008). More concrete, R. Idrisov (2013) demonstrates that among 110 units of words denoting basic concepts (100 of them taken from the Swadesh's list[2] and 10 from the supplemental list of Jakhontov[3] there are 99 units which are common for Beserman idiom and Udmurt language. For these reasons, we treat Beserman as a dialect of Udmurt.

The question of Beserman ethnogenesis is still discussed and is far from getting any satisfactory answer. For different hypotheses see Napolskikh (1997), Teplyashina (1970),

[2] The Swadesh's list (see Kassian, Starostin, Dybo & Chernov 2010) represents 100 most stable items of the basic vocabulary of human languages. After certain modifications, it has become a very important tool of comparative linguistics.

[3] S. Yakhontov has offered a modified version of Swadesh's list and divided it into the most stable part (35 words) and the part that is lost more quickly (65 words). A common praxis of lexicostatistical studies now is to combine the Swadesh's list of 100 words with the following 10 cognates offered by Yakhontov: far, heavy, near, salt, short, snake, thin, wind, worm, year. For the reasons and details see Burlak & Starostin (2005: 12-13).

Nasipov (2010). But it is indisputable that the history of the ethnic group is rich in contacts with speakers of different languages from several language families. Some of these languages – primarily Tatar, Udmurt and Russian – still keep their important role in formation of the idiom in question. Here we will focus on the influence which Russian exerts on oral speech of Beserman speakers living in Shamardan (Yukamenskoye district, Udmurtia, Russian Federation).

All Beserman speakers living in Shamardan are multilingual. They speak at least Beserman dialect which they strictly separate from Udmurt literary language, Udmurt and Russian. Beserman is the means of every-day communication. Udmurt is acquired at primary school. It is used while reading school textbooks and regional newspapers and in talks with Udmurt neighbors from the same village. Russian is learned at school. It is the official language of the country, i.e. the language of all situations except every-day communication: it is used to speak with officials, teachers at secondary school, while watching TV, reading most newspapers and books etc. Being the main means of communication in most situations, Russian has a great influence on the speech of Beserman people. The easiest way to analyze this influence is to look at a corpus of texts which contains transcriptions of Beserman oral speech. To make such analysis is the main aim of this paper.

It should be noted that the problem of the influence of Russian on Udmurt has already attracted the researchers' attention. Different aspects of it from Udmurt-Russian code-switching and code-mixing to adoption of Russian grammatical features in Udmurt have been studied. For the most recent results in this field see, for example, Solomennikova (2012), Shirobokova (2011), Kaysina (2013). However, as far as we know, the Beserman dialect has not been taken into account from this point of view yet.

We start by defining the object of our study. We investigate the influence of Russian on Beserman by analyzing the cases of code-mixing and recent lexical borrowing. The terminus "code-mixing" is used here in the same sense as in Muysken (2000) referring to cases when lexical items, grammatical (and, as we must add, phonetical) features of several languages are used in one utterance. Examples of code-switching, i.e. "the rapid succession of several languages in a single speech event" (Muysken 2000: 1), as in (1), also appear in our corpus of Beserman texts:

(1) Beserman Udmurt fieldnotes:

Bakč'a-ja-z	sâl-e.	Vot	vo	sâl-e,	sâl-e,	a	ja
kitchen.garden-Loc-P.3	stand-Prs.3Sg	DM_{rus}	DM_{rus}	stand-Prs.3Sg	stand-Prs.3Sg	$Conj_{rus}$	me_{rus}

duma-l-a	uže...	e,	ani-je!	Malpa-j	n'i	gurt-a-z	pâr-i-z	šuâsa.
think-Pst-3Sg_{rus}	now_{rus}	DM	mother-P.1	think.Pst	now	house-Ill.-P.3	enter-Pst-3	Conj

She is in the kitchen garden, here, here she is, and I was sure that she was there, she was, oh, my gosh! I was sure, that she had entered the house (underlined fragment of the sentence is in Russian).

Such cases are beyond the scope of this article, but we would like to notice that examples of code-switching from our corpus seem not to be motivated by any external factors like change of the speech situation, of the addressee etc. So, the situation with Beserman, Russian and Udmurt, at least in Shamardan, is not that of "ideal multilinguism" (Weinreich 1963). But nevertheless the language competence of the speakers elder than 25 years old seems to be very good: they speak with each other fluently, the cases of switching to Russian are relatively rare and the length of Russian fragments of discourse normally does not exceed 1-2 sentences (at least in informal talks).

We must stress that we distinguish code-mixing from lexical borrowing. It is not a trivial task since there is a considerable amount of Beserman roots loaned from Russian. Thus, R. Idrisov (2013) shows that in the sample of 1912 non-derived Beserman roots 48% are common Permic, 14% loaned from Turkic languages, 17% loaned from Russian and 21% do not have reliable etymologies (Idrisov 2013). It is a common praxis to treat as borrowed words those which are included in normative dictionaries of the investigated idioms (for Udmurt this strategy was used in, for example, Solomennikova (2012)). But the on-line dictionary of the Beserman dialect (http://beserman.ru) is still under elaboration. So we treat as loanwords those lexical units which are phonetically and morphologically adopted. All the other cases when Russian words appear in our corpus we consider to be the cases of code-mixing. For example, the lexical unit *petu**k*** 'rooster' is loaned from Russian, but it was phonetically transformed according to Beserman phonetic rules, whereas *petu**x*** 'rooster' is a Russian word with Russian phonetic structure

(Beserman system of consonants does not include the sound [x] (Idrisov 2013: 33-34)), so its appearance in a Beserman text is a case of code-mixing.

As for grammar features, we also (along with code-mixing) take into account the cases of loaning a Russian grammatical unit in order to cover a gap in Beserman system.

After defining the object of our investigation let us describe the corpus of texts we used. We shall try to show how the Russian influence is manifested in 19 texts (about 2,5 hours of oral speech) recorded and transcribed in Shamardan in January and July 2010. The texts were collected during experiments concerning conditions of the choice between local cases and postpositions for expressing spatial relations. There are three groups of texts corresponding with the tree types of experiments: 10 monologues describing the events developing in a 4-minute cartoon; 6 dialogues where the main speaker explains the second speaker how to place cards on a picture; 3 dialogues containing explanations how to move three figures through a model of the Shamardan area.

Cartoons, cards, figures, locations and roots were identical for all the speakers. The texts were recorded from 11 speakers (6 men and 5 women). Two speakers were young women between 25 and 30 years old, the rest were elder than 60. The reason is that there are only a few Besermans under 30 living in Shamardan permanently (it is a pity because young Besermans in general speak quite fluently). As for children, they do not acquire Beserman as a mother tongue being able only to understand Beserman oral speech and using some words. In the given 19 texts Russian influence can be retraced at practically all the levels of the idiom: phonetics, morphology, vocabulary, syntax, and discourse.

2. Phonetics

In Beserman there is a strong tendency to conform to the Beserman phonetic rules in speech. But nevertheless the Russian influence exists even at this level. We will illustrate it with two examples.

First, as mentioned before, in Beserman system there is no sound [x], but it may occur in Russian words used in fluent speech instead of phonetically adopted Beserman ones:

(2) Beserman Udmurt fieldnotes:

Kwin',	n'ul',	vit'-et'i-jez	petu**x**.
three	four	five-Ord-P.3	$rooster_{rus}$

[counting the figures] *Three, four, and the fifth is a rooster.*

Such pronunciation of the word denoting 'rooster' is treated by the native speakers we have worked with as being Russian (not Beserman). In their judgment, the Beserman word for 'rooster' is [petuk], as in (3):

(3) Beserman Udmurt fieldnotes:

Petu**k**	no	kureg	no	soos	mân-o	vič'ak	š'iš'kâ-nâ.
$rooster_{rus}$	and	hen	and	they	go-Prs.3Pl	everybody	eat-Inf

The rooster and the hens, everybody is going for eating.

Second, there is also influence in suprasegmental domain. Udmurt stress is fixed on the final syllable of the word. N. Ljukina mentions that Russian words are often loaned in Beserman with conserving of Russian stress (Ljukina 2007: 134). According to our data, there is a tendency (at least in Shamardan) to pronounce loanwords with Beserman stress. But in fluent speech some Russian words with Russian stress may occur, though not very often. Compare:

(4) Beserman Udmurt fieldnotes:

Tatân...	tat'jân		ik	pič'i	pi-len	kol'**a**ska-jez,	pič'i	veloš'iped-ez.
here	here	[correct: tatân]	DM	little	boy-Gen1	$pram_{rus}$-P.3	little	$bicycle_{rus}$-P.3

There is a pram of a little boy, a little bicycle here.

Dialectologia 17 *(2016), 123-150.*
ISSN: 2013-2247

3. Morphology

From now on we will take into account two phenomena on each language level examined – code-mixing and borrowing.

3.1 Code-mixing

In our corpus occur the following cases of using Russian morphological tools instead of Beserman ones. Thus, instead of Beserman *inchoative* construction "*kuč'kə̂nə̂* 'to start' + infinitive" appears the Russian construction "*davaj* 'let' + infinitive":

(5) Beserman Udmurt fieldnotes:

I	davaj	až'-lan'	mânə̂-nə̂.
and	$Inch_{rus}$	forward-All	go-Inf

And they began to go forward.

Constructions with inchoative meaning also represent a very interesting type of code-mixing which appears in our corpus several times involving different morphological units. Namely, Russian means of expressing a grammatical meaning is sometimes used together with (or contaminated with) its Beserman analogue. In such cases the meaning is actually expressed two times by morphological units from different idioms. Example (6) represents this peculiar "double-marking" of inchoative meaning, examples (9) and (10) below – of optative and debitative meanings correspondingly. "Double-marking" appears to be the most common strategy of borrowing which can be found on all language levels except phonetics (more examples will be presented later).

(6) Beserman Udmurt fieldnotes:

L'eg'it'...l'eg'it'		pi	pešt-i-z,	i	davaj	žug'iš'kə̂-nə	kuč'k'-i-z-ə̂.
young	young	man	fall.down-Pst-3	and	$Inch_{rus}$	fight-Inf	begin-Pst-3-Pl

The young man fell down, and (they) began to fight.

There are also cases of using a Russian *negation* marker instead of a Beserman one:

(7) Beserman Udmurt fieldnotes:

Petuk	n'e	so.
$rooster_{rus}$	Neg_{rus}	this

This is not a rooster.

This case of morphological code-mixing occurs in the corpus occasionally. The speakers strongly tend to use the Beserman marker. Example (8) belongs to the same speaker as (7), and it appeared in the same dialog:

(8) Beserman Udmurt fieldnotes:

Evəl	eta	petux.
Neg	$this_{rus}$	$rooster_{rus}$

This is not a rooster.

Next, *optative* can be expressed through contamination of Beserman and Russian morphological tools:

(9) Beserman Udmurt fieldnotes:

Mar	ke	otân...	aj,	olo	kâž'ə vâldə...	ja	goroč'ka,	puskaj	med	lu-o-z...
that	Cond	there	DM_{rus}	or	how DM	DM	$little.hill_{rus}$	OPT_{rus}	OPT	be-Fut-3

u-g	tod-iš'k-ə	kâž'ə	vera-nə.
Neg.NPst-3	know-Prs-Sg	how	tell-Inf

There is something... auch, or as it is known... well, let it be a little hill, I don't know, how to put it.

Compare:

(9a) Russian:

Puskaj	bud'-et.
Opt	be.Fut-3Sg

Let it be.

Example with "double-marking" of *debitative* meaning:

(10) Beserman Udmurt fieldnotes:

I	van'	tânad	otân	lu-o-z	veloš'iped,	vel'ik	otân	tânad	dolžen lu-ânо.
and	is	you.Gen1	there	must.be-Fut-3	$bicycle_{rus}$	$bike_{rus}$	there	you.Gen1	$must_{rus}$ must.be-Deb

And you must have in there a bicycle, you must have a bike in there.

2.2 Borrowing

There are also several cases of loaning morphological units from Russian in order to cover the gaps in Beserman system. One of them is using the Russian word *davaj* to form *imperative of the first person* which does not exist in Beserman (Teplyashina 1970: 236):

(11) Beserman Udmurt fieldnotes:

Soldat	vera	Ivan-lâ: "Davaj	mon	pe	tâb-o	kâz	jâl-e	aǯ'-o,
$soldier_{rus}$	say.Pst.3Sg	Ivan-Dat	$Imp.1_{rus}$	I	Quot	climb-Fut.1Sg	firtree	top-Ill	see-Fut.1Sg

mar	ke	u-g	a	aǯ'iš'kâ.
that	Cond	Neg.NPst-3	Q	be.seen

The soldier said to Ivan: "Let me climb the firtree top, maybe something is seen (from there)".

(12) Beserman Udmurt fieldnotes:

Maša,	davaj	aš'mes	kuč'k-o-m	skal-jos	dor-iš'en.
Masha	$Imp.1_{rus}$	Refl.1Pl	start-Fut-1Pl	cow-Pl	neighbourhood-Egr

Masha, let us start with the cows.

The other case is the Russian word *samoj* which function in Beserman as *superlative* marker (in Beserman there is no other means to express this meaning) (Teplyashina 1970: 179):

(13) Beserman Udmurt fieldnotes:

samoj	umoj
$Superlat_{rus}$	good

the best

This marker can be joined also to words functioning as nouns:

(14) Beserman Udmurt fieldnotes:

Velošʼiped	wanʼ	vâl-ân,	samoj	vâl-ân.
bicycle$_{rus}$	there.is	top-Loc	Superlat$_{rus}$	top-Loc

There is a bicycle above, at the top.

4. Vocabulary

4.1 Code-mixing

Such cases are relatively rare, but they do exist:

(15) Beserman Udmurt fieldnotes:

Tabere	Paša	L'emskoj-ən	opetʼ,	korka	šʼer-a-z,	d**a**lʼnʼij	korka	šʼer-a-z.
then	Pasha	L'emskij-Loc	again$_{rus}$	house	behind-Loc-P.3	distant$_{rus}$house		behind-Loc-P.3

And then, Pasha is now in Lemskij (a settlement near Shamardan) *again, behind the house, behind the distant house*

(note that the Russian word *dalʼnʼij* 'distant' is used instead of the Beserman widespread word *kâďokâšʼ* 'distant'; bold type indicates stress).

4.2 Borrowing

In our corpus of texts two cases of borrowing the vocabulary units occurred. In the first case, the new word is coming together with the new actual. With the lapse of some time, such words are phonetically and morphologically adopted:

(16) Beserman Udmurt fieldnotes:

I	majeg	jâl-a-z	sâl-e	puktâ-mân	čʼâgân.
and	stake	top-Loc-P.3Sg	stand-Prs.3Sg	put-Res	cast-iron.pot$_{rus}$

And somebody has put a cast-iron pot at the top of the stake, and now it is there

(in Russian the word denoting 'cast-iron pot' is pronounced as [chugun]).

(17) Beserman Udmurt fieldnotes:

Pios	murt	lâkt-e	traktor	dor-e	traktor-z-e	zavod'-tâ-nâ.
man	human	come.back-Prs.3Sg	$tractor_{rus}$	neighbourhod-Ill	$tractor_{rus}$-P.3-Acc	start.up-Tr-Inf

A man is coming back to a tractor to start it up

(compare Russian *zavod'i-t'* 'to run up' where *-t'* is an infinitive suffix).

In the second case, the Russian word (chashja 'a thick forest' is a Slavic word, see Vasmer 1986-1987) replaces the Beserman one. In this case the word also can be phonetically (18) or/and morphologically (19) adopted:

(18) Beserman Udmurt fieldnotes:

Pič'i	pi	vel't'â-mân	č'ašja-je.
little	boy	walk-Res	$forest\text{-}Ill_{rus}$

Little boy went to the forest

(compare Russian *chash'a* 'a thick forest').

(19) Beserman Udmurt fieldnotes:

Odig-âz	pâd-âz	gol'ik,	odig-âz	noski-jen.
one-P.3	leg-P.3	bare	one-P.3	$sock_{rus}$ -Instr/Comit

One leg is bare and the other is in the sock

(the Russian word *noski* is a plural form of a word *nosok* 'sock'; note also that there is an old Beserman word *pâd nâr* 'sock, socks').

It is a productive model in Beserman to adopt the frequent plural forms of Russian nouns as words denoting one object. Here are two more examples of this kind:

(20) Beserman Udmurt fieldnotes:

piroški-os

pie-Pl

pies (The Russian word *piroshk'i* is a plural form and already denotes many pies.)

(21) Beserman Udmurt fieldnotes:

štan

trousers

one pair of trousers

(The Russian word *shtany* is *pluralia tantum*; in Beserman *-y* is lost in Nominative, but it still occurs as *-i* in the oblique stem of the noun.)

In the case of replacing a Beserman word through a Russian one also a semantic shift may occur. A good example is the word *sad*. In Russian it denotes a plot planted with trees, bushes and flowers or trees and bushes growing on such plot (Dal 1978). In Beserman this word has replaced the word *č'ešpel'* 'bushes, small young trees' and has acquired a meaning 'leaf-bearing tree' (having in this meaning also a plural form). (As far as we know, there is no Beserman word expressing the concept of a leaf-bearing tree – as there is no Beserman word denoting the coniferous tree).

Here we must say a few words about the difference among words belonging to different morphological classes with respect to the degree of adaptation. First of all, *pronouns* are never adopted. *Numerals* are adopted extremely seldom. Thus, we know only one numeral adopted from Russian - *pervoj* 'first'. For *postpositions* we also know only one case of substitution by a Russian word. Namely, the Beserman postposition *mestaje* (*mesta*-Ill) 'instead of' is a loan translation of the Russian preposition *vm'esto* (*v-mesto* = in-place.Acc) 'instead of'. *Adverbs* are adopted rarely, and their degree of phonetic adaptation is close to that of nouns which may either be adopted (consider the word *opet'* 'again' in (15) which in Russian is pronounced as [op'at']; this word has completely substituted an old Beserman word *nəš* 'again') or not (22):

(22) Beserman Udmurt fieldnotes:

Obratno	tatč'ə	lə̂kt-i-z.
backwards$_{rus}$	here	return-Pst-3

[She] returned here, backwards

(Note that there is a widespread Beserman adverb *berlan'* 'backwards'.)

Nouns are adopted most often, and they may be phonetically or/and semantically adopted or not (see (16) and (18) above). There are also cases of grammatical adaptation of nouns, most of them concerning the number of the noun (see (19)-(21) above).

Adjectives are often adopted, and they may also demonstrate phonetic and semantic shift or not (see (15) again).

Verbs are generally adopted without semantic shift but they are very good adopted phonetically. Furthermore, Beserman demonstrates two productive models of verbal adaptation. First, some verbs are adopted through taking the verbal stem (phonetically adopted, if necessary) and attaching Beserman inflexions – sometimes even derivational suffixes – to it. The second (and much more universal) strategy is to take the infinitive of the Russian verb and to attach the inflectional forms of the Beserman verb *karə̑nə̑* 'to do' to it. To use such a strategy is often the first reaction during the speech generation even if a good Beserman verbal equivalent exists. But while speaking to us Besermans often corrected themselves at once:

(23) Beserman Udmurt fieldnotes:

Soldat	soje	opat'	spaš't'i	kar-e,	e,	jurt-[e]
soldier	he.Acc	again$_{rus}$	to.save$_{rus}$	do-Prs.3Sg	Autocorr	save-[Prs.3Sg]

n'e	jurt-e,	n'e spasajet,	a	jurt-e.
Neg$_{rus}$	save-Prs.3Sg	[not saving]$_{rus}$	but	save-Prs.3Sg

The soldier again saves him…no, saves, not "spasajet" but "jurte".

Conjunctions are adopted very actively replacing the Beserman ones. In general they are adopted neither phonetically nor functionally. Compare:

(24) Beserman Udmurt fieldnotes:

Ben,	so	daže	n'e	č'už-g'ez,	a	kə̑č'eke	go[rd]...	marə̑m...
yes	it	even$_{rus}$	not$_{rus}$	yellow-Compar	but$_{rus}$	some	red	maybe

Yes, it is even not yellow but red or the like... maybe…

Finally, during the borrowing of words the "double marking" of sense may occur. Thus, a Russian word expressing a concept may occur together with a Beserman one which has the concept in its semantic structure as a component. Compare:

(25) Beserman Udmurt fieldnotes:

Mə̑nam	even'	bol'she	kureg-jos-ə̑.
I.Gen1	have.no.anymore	$more_{rus}$	hen-Pl-P.1

I have no hens anymore.

It is notable that the order can be reversed (in Beserman there is a very small amount of expressions where the order of constituents can be reversed most of them being onomatopoetic ones):

(26) Beserman Udmurt fieldnotes:

bol'she	even'
$more_{rus}$	is.not.anymore

Not anymore.

Thus, the following hierarchy of susceptibility of different morphological classes to borrowing can be built for Beserman dialect:

conjunctions < nouns < verbs < adjectives < adverbs < postpositions < numerals < pronouns.

In the case of susceptibility to borrowing Beserman is a very typical idiom. A close hierarchy (perhaps more detailed) can be drawn from (Matras 2009: 166-208).

5. Syntax

5.1 Code-mixing

According to our corpus, Russian influence on Beserman syntax was not at all as huge as on the vocabulary. There is a large amount of Beserman constructions which in

our corpus (and according to our experience) are never replaced with Russian ones. One of them, for example, is a construction "finite verb + *mə̂n*-participle" which has a resultative meaning (see (16)). But nevertheless there are some cases where Beserman constructions can be replaced by Russian ones. The first of them is the construction with sentential actants:

(27) Beserman Udmurt fieldnotes:

Aǯ'-i-z	što	nə̂l	murt	š'ud-e	odig-ez	petux-ez	evə̂l,	pič'i	pi
see-Pst-3	$Conj_{rus}$	girl	human	feed-Prs.3Sg	one-P.3	$rooster_{rus}$-P.3	not	little	boy

petux	punna	mə̂n-i-z	biž'-ə̂sa.
$rooster_{rus}$	for	go-Pst-3	run-Conv

He saw that the girl was feeding [birds] and that one chicken was missing, and the boy ran for the chicken.

Compare:

(27a) Russian:

Uv'id'e-l-Ø	što	d'evuška	korm'-it	kur-Ø.
see-Pst-M.Sg	Conj	girl	feed-Prs.3.Sg	hen-Gen.Pl

He saw that a girl was feeding hens.

(27b) illustrates the common Beserman construction with the verb *aǯ'ə̂nə̂* 'to see':

(27b) Beserman Udmurt fieldnotes:

Pi...	pič'i	p'i	lə̂kt-i-z	gibija-m-ə̂š'	korž'inka-z-e	kel't-i-z,
boy	little	boy	come.back-Pst-3	pick.mushrooms-Nzr-El	$basket_{rus}$ -3-Acc	leave-Pst-3

aǯ'-i-z-ə̂	odig-ez	kureg-ez	potə̂-mə̂n	vol'noj.
see-Pst-3-Pl	one-3	hen-P.3	come.out-Res	$free_{rus}$

The little boy came back from picking mushrooms, left his basket and saw that a hen had come out, [and now it is] free.

The second construction is the comparative one:

(28) Beserman Udmurt fieldnotes:

So	kak budto	ǯ'aǯ'eg	pi-jez-lâš'	bâž-z-e	kurč'-e	n'i.
he	$Compar_{rus}$	goose	son-P.3-Gen2	tail-P.3-Acc	bite-Prs.3Sg	Emph

It seems that he is biting the gosling's tail.

There are several types of Beserman comparative constructions. The most universal of them is the construction with *kaď* 'like':

(29) Beserman Udmurt fieldnotes:

Ben,	ulla	pal-a-z	kaď.
yes	lower	side-Loc-P.3	like

Yes, it seems to be a bit lower.

Russian and Beserman comparative words can be used together forming a construction with "double marking":

(29a) Beserman Udmurt fieldnotes:

So	tin'	pânâ	pi-jen	tuš'-en	visk-ân	kaď'	kak by.
it	well	dog	son-Instr/Comit	trough-Instr/Comit	between.Obl-Loc	like	$like_{rus}$

Well, it seems to be between a puppy and a trough.

It should also be noted that sometimes (hardly ever) the Russian verb frame is used together with the borrowed verb itself:

(30) Beserman Udmurt fieldnotes:

Tak,	fs'o,	razobra-l'-i-š'	kureg-jos-ân,	ben	a?
$well_{rus}$	$that's.all_{rus}$	$deal\text{-}Pst\text{-}Pl\text{-}Refl_{rus}$	hen-Pl-Instr/Comit	yes	Q

Well, that's all, we have finished with the hens, haven't we?

Compare:

(30a) Russian:

Razobra-l'-i-s'	s	kur-am'i.
deal-Pst-Pl-Refl	with	hen-Instr.Pl

We have finished with the hens.

The third construction is the coordinative one. The Beserman coordinative constructions with conjunction *no…no* 'and' (31) is often replaced with the Russian one with conjunction *i* 'and' (32):

(31) Beserman Udmurt fieldnotes:

Ad'ami-z-e	no	kut-i-z	no	vu	pušk-ə̂	dong-i-z.
man-P.3-Acc	and	catch-Pst-3	and	water	in.Obl-Ill	push-Pst-3

[He] caught the man and pushed [him] into the water.

(32) Beserman Udmurt fieldnotes:

Ivan,	soldat	soje	pə̂d-t'i-z	kə̂sk-e	kə̂sk-e	i	ǯ'ut-i-z	vu-ə̂š'.
Ivan	$soldier_{rus}$	he.Acc	leg-Prolat-P.3	draw-Prs.3Sg	draw-Prs.3Sg	and_{rus}	pull.out-Pst-3	water-El

(As for) Ivan, soldier is drawing and drawing (him) by his legs and pulled out of the water.

The Russian construction occurs much more often than the Beserman one. Besides, the "double marking" may also occur here:

(33) Beserman Udmurt fieldnotes:

L'eg'it'	pi-z-e	no	miš'k-e	i	jə̂...jə̂rč'ə̂-t'i-z	kutə̂-sa	vu-e	donga.
young	man-P.3-Acc	and	wash-Prs.3Sg	and_{rus}	hair-Prolat-P.3	catch-Conv	water-Ill	push.Pst.3

[He] washes the young man and after catching [him] by his hair pushes [him] into the water.

6. Discourse

6.1 Code-mixing

In use of discourse markers (DM) the general tendency of combining the Russian and the Beserman variants ("double-marking") can be seen again:

(34) Beserman Udmurt fieldnotes:

Ja	ladno,	pun-i-d	ke	pun-i-d.
well	$well_{rus}$	put-Pst-2	Cond	put-Pst-2

Ok, if you have put [it] than you have put [it].

Another strategy is to replace one of the components of a complex Russian DM with a Beserman equivalent of this component:

(35) Beserman Udmurt fieldnotes:

Nu	ten'.
well	here

Well, we've finished.

Beserman *ten'* 'here' corresponds to Russian *vot* 'here', and *nu ten'* in (35) is used just as Russian marker *nu vot*.

6.2 Borrowing

Russian DMs can be also borrowed directly:[4]

(36) Beserman Udmurt fieldnotes:

Bur-laš'an',	otân	tin'	na	marâm	eto...	ooo,	oj, bl'in...		otân	van'	na	tânad
right-Recess	there	here	Emph	that	DM_{rus}	oh	oh	DM_{rus}	there	be.Prs.3	Emph	you.Gen1

eto	dâdâk-jos-âd	van'	na.
DM_{rus}	pigeon-Pl-P.2	be.Prs.3	Emph

[4] Note that in a close-related Komi there is also a good amount of discourse words borrowed from Russian (Leinonen 2014).

From the right side, there is... hm, oh, oh, damn, there is you have... eee... there are pigeons (underlined markers are adopted from Russian).

All the DMs appeared in our corpus are presented in Table 3 and Table 4 (see Appendix). In Table 3 for each DM the following information is given: a) its function in the discourse; b) its origin (Russian, Beserman, Russian-Beserman (mixed) or unknown); c) its phonetic form; d) number of occurrences in the corpus; e) number of speakers which have used it. The number of speakers and the number of occurrences decreases from the top of the table to its bottom. Note that the functions of the markers are labeled very roughly. The studying of Beserman DM is only at an initial stage, so we can in no case pretend to present a complete analysis of them. The functional labels are given there just for illustrative purposes.

Table 4 demonstrates the correspondence between Beserman markers and their supposed Russian analogues. Again, for each marker the origin (Beserman or Russian), the function, the phonetic form, number of occurrences in the corpus and number of speakers which have used it are given.

The data given in the Appendix are summarized in Table 1:

Origin	**Degree of generality**			
	The most frequent (5-10 speakers)	Less frequent (2-4 speakers)	Rare (1 speaker)	All
Russian	11	18	34	63
Beserman	17	10	8	35
Mixed (“double-marking”)	0	4	0	4
Unknown	2	11	8	21

Table 1. The distribution of DM of different origin used in Beserman dialogs according to their degree of generality

Here the “degree of generality” means number of speakers who used DMs of given origin. DMs which appeared as a result of the “double-marking” strategy are labeled as “mixed”. Calques from Russian are concerned to have Russian origin.

One can see from Table 1 that the amount of DMs which are supposed to be borrowed from Russian tend to increase with the decreasing of the amount of people using it. Roughly speaking, we can say that Russian DMs tend to be individual (except *nu* which has practically substituted the corresponding Beserman marker *ma*). The amount of markers with Beserman origin, on the contrary, tend to decrease with the decreasing of the amount of speakers – roughly speaking, they tend to be more common. On the other hand, the total amount of Russian DMs significantly exceeds the total amount of Beserman ones (63 vs 35 respectively). The entire situation looks like that the influence of Russian on Beserman has significantly grown recently: different speakers have acquired different DMs which have not stabilized yet in the language praxis of the community. DMs and connectors are claimed to be the group which is borrowed more quickly than other grammatical function words (Matras 2009) and than many other common linguistic items (Leinonen 2014), so the beginning of the active borrowing of them seems to be a good criterion of a new wave of significant influence from an external idiom.

7. Conclusions

Results of the analysis presented in this paper are summarized in Table 2:

Type of influence	Language level				
	Phonetics	**Morphology**	**Vocabulary**	**Syntax**	**Discourse (DMs)**
Code-mixing	Words with Russian phonetic structure and/or Russian stress	Inchoative Optative Debitative Negation	Rare	Comparative constructions Constructions with the sentential actants Coordinative construction	Not very often
Borrowing	----------	Imperative of the 1st person Superlative	conjunctions < nouns < verbs < adjectives < adverbs < postpositions < numerals < pronouns	Valency patterns (extremely rare)	Very often

Table 2. Russian influence on Beserman Udmurt on different language levels

Observing all cases of Beserman-Russian code-mixing and of borrowing from Russian into Beserman which are presented in our corpus of texts we can see that the degree of influence is not as great as it might seem to be. A really great amount of units borrowed from Russian essentially appears only among nouns, but such loanwords tend to be phonetically, morphologically and even semantically adopted. Loanwords belonging to other parts of speech either occur seldom or are phonologically adopted. As for morphology, there are not many tools borrowed, most of them fulfilling the gaps in the Beserman system. The most frequent discourse markers in general have Beserman origin. The syntax practically has not been influenced by Russian. In the discourse of 25-30 years old speakers there are more adopted constructions and morphological borrowed units than in the discourse of 60-85 year old people, but the disparity is not very big. A possible reason is that young Besermans use their mother tongue while talking to each other and to their relatives.

The degree and peculiar properties of Russian influence on Beserman dialect lead to the suggestion that in more favorable sociolinguistic situation the result could be a mixed idiom with predominantly Russian vocabulary and Beserman basis of grammar. The loaned language units are likely not to be copied without noticeable changes, but to be adopted, even in special strategies of borrowing ("*karânâ* 'to make' + infinitive", "double marking"). Examples of mixed languages are numerous (for one of the most recent of them, Gurundji Kriol, which arose in the 1970s- 80s, see McConvell & Meakins 2005). Mixing with other idioms (or, in other terms, copying vocabulary units and grammar features of them (Johanson 2002) seems to be a normal phenomenon in a language's life. Linguists have even started to take it into account while building family trees of languages (McMahon & McMahon 2006). But Beserman dialect unfortunately has no chance to reach this stage. It is supposed to die in recent years because people under 25 do not speak Beserman. All the grandchildren and little children of people we worked with in Shamardan do not speak Beserman. Most of them even do not understand Beserman speech. The idiom is also not taught at school. Thus, Beserman Udmurt seems to be a sad example confirming Lars Johanson's statement that codes never die of structural changes

of any kind, they "die because they are no longer acquired by the new generations" (Johanson 2002: 267).

References

BURLAK, Svetlana A. & Sergey A. STAROSTIN (2005) *Sravnitelno-istoricheskoye yazykoznaniye* [Comparative Linguistics], 2nd edition, Moscow: Akademiya.

DAL, Vladimir I. (1978) *[Tolkovyj slovar zhivogo velikorusskogo jazyka]*, v. 1-4, Moskva.

IDRISOV, Ruslan I. (2013) *Turkish Loanwords of Different Periods in Beserman Udmurt* [Tjurkskije zaimstvovanija raznykh periodov v besermianskom dialekte udmurtskogo jazyka], Thesis, Moscow, Manuscript.

JOHANSON, Lars (2002) "Do languages die of 'structuritis'? On the role of code-copying in language endangerment", *Rivista di Linguistica*, 14(2), 249-270.

KAYSINA, Inna (2013) "The adoption of Russian Conjunctions in Udmurt", ESUKA-JEFUL 2013, 4-2, 131-144.

KASSIAN, Alexei, George STAROSTIN, Anna DYBO & Vasiliy CHERNOV (2010) "The Swadesh wordlist. An attempt at semantic specification", *Journal of Language Relationship*, 4, 46-89.

KELMAKOV, Valej K. (1998) "A brief sketch of Udmurt dialectology. Introduction. Phonetics. Morphology. Texts on Udmurt dialects. References" [Kratkij kurs udmurtskoj dialektologii. Vvedenije. Fonetika. Morfologija. Dialektnyje teksty. Bibliografija], Izhevsk: Udmurt State University.

LEINONEN, Marja (2014) "Borrowability of function and discourse words: the Komi case", *Language Contacts: the State of the Art*, Helsinki, August 28-30, 2014, Report.

LJUKINA, Nadezhda M. (2007) "Word Stress in Beserman Patois of Udmurt [Slovesnoje udarenije v besermianskikh govorakh udmurtskogo jazyka]", *Current Problems of Science and Education* [Sovremennyje problemy nauki i obrazovanija], 6, Moscow: Akademija obrazovanija, 132-135.

LJUKINA, Nadezhda M. (2008) *The peculiarities of the language of Balezino and Jukamenskoje Besermans*, PhD-thesis. [Osobennosti jazyka balezinskikh i jukamenskikh besermian.] Moscow, Manuscript.

MATRAS, Yaron (2009) *Language Contact*, Cambridge: Cambridge University Press.

MCCONVELL, Patrick & Felicity MEAKINS (2005) "Gurindji Kriol: A Mixed Language Emerges from Code-switching", *Australian Journal of Linguistics*, 25(1), 9-30.

MCMAHON, April & Robert MCMAHON (2006) "Keeping Contact in the Family: Approaches to Language Classification and Contact-induced Change", in Yaron Matras *et al.* (eds.), *Linguistic Areas. Convergence in Historical and Typological Perspective*, Houndmills: Palgrave Macmillan, 51-74.

MUYSKEN, Pieter (2000) *Bilingual Speech: A Typology of Code-Mixing*, Cambridge: Cambridge University Press.

NAPOLSKIKH, Vladimir V. (1997) *Introduction to Historical Uralistics* [Vvedenije v istoricheskuju uralistiku], Izhevsk: Udmurt Institute of History, Language and Literature, Ural Branch of Russian Academy of Sciences Publ.

NASIPOV, Ilshat S. (2010) *Finno-Ugric Loanwords in Tatar: Synopsis and Taxonomy* [Finno-ugorskije zaimstvovanija v tatarskom jazyke: sinopsis i taksonomija], Doctoral thesis, Kazan, Manuscript.

SHIROBOKOVA, Larisa (2011) *Udmurt-Russian Bilingualism (Udmurt Republic, Sarkan Region, Muvyr Village)*, PhD-dissertation, Budapest, Manuscript.

SOLOMENNIKOVA, Anastasija P. (2012) "Code-switching and Code-mixing in the Udmurt Speech of Udmurt-Russian Bilinguals (fieldwork experiment)" [Perekljuchenije i smeshenije koda v rechi udmurtsko-russkikh bilingvov (opyt polevogo issledovanija)], *The Questions of Philology* [Voprosy filologii], 3 (42), 89-98.

TEPLYASHINA, Tamara I. (1970) *The Language of Beserman* [Jazyk Besermian], Moscow: Nauka.

VASMER, Max (1986-1987) *Russian Etimological Dictionary* [Etimologicheskij slovarj russkogo jazyka], 4 volumes, Moscow: Progress.

WEINREICH, Uriel (1963) *Languages in Contact: Findings and Problems*, 2nd edition, The Hague: Mouton.

Appendix

Frequency of DMs in the corpus

function	Beserman / Russian	marker	corpus frequency	number of speakers
emphasis	bes	n'i	182	10
question	bes	a	162	10
emphasis	rus	nu	113	10
emphasis	bes	na	72	10
agreement	bes	ben	264	9
emphasis	bes	uk	95	8
end/beginning of the episode	rus	tak	90	8
beginning of the new episode	rus	a	76	8
request for attention	bes	ik	59	8
declining the speaker's responsibility for the content of the utterance	bes	mar a	34	8
perplexity	rus	eto(t)	21	8
change of topic	bes	sə̂re\so bere	60	7
verification	bes	ben a	49	7
agreement against the wish of the speaker	bes	ja	37	7
understanding	rus	a	21	7
perplexity	rus	naverno	14	7
autocorrection	bes	e	11	7
agreement	rus	aha	62	6
end of the episode	bes	tin'	45	6
emphasis	bes	ma	18	6
request for some time	rus	š'a(s)	14	6
end of the episode	rus	fs'o	13	6
agreement	?	mh	81	5
request for some time	bes	moga	38	5
end of the episode	rus	vot	36	5
quotation	bes	pe	28	5
hesitation	?	eeee	25	5
[yet failed to be determined exactly]	rus	eššo	19	5
negation	bes	evə̂l	13	5
declining the speaker's responsibility for the content of the utterance	bes	mar ke	10	5
change of topic	bes	ta(be)re	37	4
hesitation	rus/bes	marə̂m eto(t)	33	4
agreement against the wish of the speaker	rus	ladno	21	4
declining the speaker's responsibility for the content of the utterance	bes	marə̂m (marə̂me)	19	4
[yet failed to be determined exactly]	rus	tože	15	4
hesitation	?	mmm, emmm	14	4

feedback	rus	a?	11	4
backchannel	rus	a	10	4
agreement	rus	da	10	4
beginning of the new episode	rus	i	8	4
error report	?	oj	7	4
agreement against the wish of the speaker	rus/bes	ja ladno	6	4
change of topic	rus	dal'še	6	4
[yet failed to be determined exactly]	?bes	no	34	3
agreement	?	m-m	12	3
verification	rus/bes	ben vet'	10	3
agreement against the wish of the speaker	rus	puskaj	8	3
perplexity	bes	leš'a	7	3
agreement against the wish of the speaker	rus	nu ladno	5	3
end of the episode	bes	ten'	5	3
perplexity	?	a	4	3
problems of speech production	?	oh	3	3
perplexity	bes	vâldâ	3	3
approximation	rus	poč't'i (što)	19	2
verification	bes	valad-a	7	2
disagreement	?	m-m	7	2
surprise	?	o	5	2
perplexity	rus	možet	5	2
hesitation	?	mh	5	2
hesitation	bes	â	5	2
declining the speaker's responsibility for the content of the utterance	rus	kak by	4	2
word search	rus	nu	4	2
autocorrection	?	fu	4	2
backchannel	bes	valaj	3	2
[yet failed to be determined exactly]	rus	v obš'em (-to)	3	2
request for some time	rus	sejčas	2	2
recomposing	rus	koroč'e	2	2
perplexity	?	o	2	2
error report	?	aj	2	2
end of the episode	rus/bes	nu ten'	2	2
autocorrection	rus	to jest'	2	2
[yet failed to be determined exactly]	bes	šu	2	2
[yet failed to be determined exactly]	rus	vet'	2	2

Besides in our corpus of texts occur a lot of markers, which are used only by one speaker. They are:

Beserman origin: *mare*, *tin' n'a (tin' na)*, *valamon*, *e*, *anije!*, *todiš'kod a, ə̂zem*, *marə̂m a*, *ož' a.*

Russian origin: *tak vet', a, a č'o, a evə̂l, bl'in, e, n'et, fs'o ravno, gad, imenno, kak budto, kak raz, kon'ešno, n'et, nu puskaj, nu tin', nu vot, podi, podoždi, pogodi, po-mojemu, primerno, razve, t'eper', t'ipa, tak-to, ten' tak, tol'ko, vern'eje, vet' tak vet', vidat', vo!, vrode, znač'it*

Unknown origin: *e, hm, m, m-m, o, t'fu, uh, uhm.*

Dialectologia 17 (2016), 123-150.
ISSN: 2013-2247

Amounts of DMs with different functions in the corpus

function	Beserman			Russian			mixed			unknown		
	marker	corpus frequency	number of speakers	marker	corpus frequency	number of speakers	marker	corpus frequency	number of speakers	marker	corpus frequency	number of speakers
agreement against the wish of the speaker	ja	37	7	puskaj	8	3	ja ladno	6	4			
				nu puskaj	1	1						
				nu ladno	5	3						
				ladno	21	4						
TOTAL		**37**			**35**			**6**				
declining the speaker's responsibility for the content of the utterance	marke	10	5	vrode	2	1						
	marem a	1	1	kak by	4	2						
	marəm (marəm e)	19	4	kak budto	5	1						
	mar a	34	8									
TOTAL		**64**			**11**							
verification	valad-a	7	2	vet' tak vet'	1	1	ben vet'	10	3			
	todiš'ko d a	2	1	tak vet'	22	1						
	ož' a	1	1	razve	1	1						
	əzem	2	1									
	ben a	49	7									
TOTAL		**61**			**24**			**10**				
request for some time	moga	38	5	sejčas	2	2				m	1	1
				š'a(s)	14	6						
				pogoďi	1	1						
				podožďi	1	1						
TOTAL		**38**			**18**						**1**	
perplexity	vəldə	3	3	t'ipa	1	1				o	2	2
	leš'a	7	3	poďi	1	1				a	4	3
				naverno	14	7						
				n'et	1	1						
				eto(t)	21	8						
				a č'o	1	1						
				vidat'	1	1						
				možet	5	2						
TOTAL		**10**			**45**						**6**	
end of the episode	tin'	45	6	vot	36	5	nu tin'	1	1			
	ten'	5	3	nu vot	1	1	nu ten'	2	2			
				fs'o	13	6						
TOTAL		**50**			**50**			**3**				
emphasis	uk	95	8	nu	113	10						
	na	72	10	kak raz	2	1						
	n'i	182	10	imenno	1	1						
	ma	18	6									
TOTAL		**367**			**116**							

change of topic	ta(be)re	37	4	t'eper'	2	1							
	səre\ sobere	60	7	dal'še	6	4							
TOTAL		**97**			**8**								
autocorrection	e	11	7	vern'eje	2	1	a evəl	1	1	t'fu	1	1
				to jest'	2	2				fu	4	2
				n'et	2	1				e, n'et	1	1
				gad	1	1						
TOTAL		**11**			**7**			**1**			**6**	
agreement	ben	264	9	kon'ešno	1	1				uhu	1	1
				da	10	4				m-m	12	3
				aha	62	6				mh	81	5
										m	3	1
										uhm	1	1
TOTAL		**264**			**73**						**98**	

Dialectologia 17 *(2016), 151-166.*
ISSN: 2013-2247

Received 17 September 2014.
Accepted 11 January 2015.

THE REGISTER OF GAMES:
A STUDY OF THE LANGUAGES OF THE GAMES IN INDIA

Amitabh VIKRAM DWIVEDI

Shri Mata Vaishno Devi University*

amitabhvikram@yahoo.co.in

Abstract

This paper is a study of the register (a variety of a language used for a particular purpose for a particular setting) of various games commonly played by children in India. The study will take into account the language of six games (three played by girls and three by boys in teenage years) for linguistic analysis in the study. The registers of the games include rhyme-scheme for the specific game, and the words associated with the games. The two genders quite vary in their languages of games. The study will find out the reflection of their sexes in the registers employed by them.

This socio-linguistic analysis will analyze the field data by examining speech practices associated with gender, and studying self-disclosure, and politeness theory in these games. The data will be observed in Hindi language (one of the official languages of India spoken in Northern India) and transcribed in International Phonetic Alphabets along with liner translation in English.

Keywords

Games, register, linguistic analysis, self-disclosure, politeness theory

* Department of Languages & Literature, Shri Mata Vaishno Devi University, Kakrayal, Jammu & Kashmir, INDIA 182 320.

EL REGISTRO DE LOS JUEGOS:

UN ESTUDIO SOBRE LA LENGUA UTILIZADA EN LOS JUEGOS EN LA INDIA

Resumen

Este trabajo es un estudio sobre el registro (una variedad de la lengua utilizada para un propósito particular en un contexto particular) utilizado en varios juegos jugados comúnmente por niños en la India. El estudio se analizará la lengua usada en seis juegos (tres jugados por niñas y tres por niños adolescentes). Los registros de los juegos incluyen el esquema rítmico para un juego específico y las palabras asociadas a los juegos. El género es una variable relevante en el estudio lingüístico de los juegos. Este estudio pretende describir cómo el género condiciona el registro utilizado en los juegos.

Este análisis sociolingüístico analizará los datos obtenidos mediante el examen de muestras de habla asociadas con el género, y estudiará, en los juegos, los mecanismos de autorevelación personal y teoría de la cortesía que se usan. Los datos se han recogido en lengua hindi (una de las lenguas oficiales de la India hablada en el norte) y se transcribirán fonéticamente junto con la correspondiente traducción al inglés.

Palabras clave

juego, registro, análisis lingüístico, autorevelación personal, teoría de la cortesía

1. Introduction

This paper is a study of the language of games commonly played by children in India. The paper will take into account the registers of various games played by both the genders. The emphasis is on children's verbal expressions in Hindi language, including rhyme-schemes, word association and selection, and other linguistic devises employed by children during the play. The focus of this work is to study the registers employed by children. This socio-linguistic analysis will analyze the data through employing ethnographic approach by examining speech practices and self-disclosure associated in these games. The data for this paper has been observed in Hindi language (an official language in India), and it has been written using Roman alphabet along with liner translation in English.

This paper finds its motivation from UNESCO's initiative to empower girls, women, and children through physical education and sport. UNESCO is presently creating an International Network on Traditional Sports and Games (TSG), and this database will be

set up with information about the main actors in traditional sports and games across the globe. The objective of this paper is to promote Indian traditional games while discussing the linguistic issues in them.

The remainder of this paper is organized under the following headings: *Value of Traditional Games* tells about the significance of traditional games in India, the heading *Indian Games* recounts some of the popular games of India, under the *Theoretical foundation* a brief account of previous studies conducted in games is presented, in the next section *Data and discussion*, the data is discussed employing cultural-historical perspective; it is supported by the content analysis, and finally the heading *Conclusion* concludes the paper.

2. Value of Traditional Games

When we watch today's children playing football, volleyball, badminton, tennis, cricket, hockey, video games, *etc.* we can trace a difference between the two generations of India-the previous generation (in 1980s and before) used to play traditional Indian games while the present generation is forgetting them. Nobody is playing games, such as *Vish-Amrit* ('Poison-Nectar'), *Kabaddi, Gulli Danda* ('Bail-Stick'), *Asta-Changa-Pay* ('Eight-Four-One'), *Kanche* ('Marbles'), *Maar-Dhaar* ('Hit and Run'), *Satolia* ('Seven Round Stones') and other traditional games in India. These games used to pass culture, heritage and traditional knowledge from one generation to another. The traditional games were designed not only to strengthen hand and eye coordination but they were structured in such a manner to assist learning which is required for multi-dimensional development, such as concentration, strategic building, logical thinking, mutual coordination, aiming, and other physical involvements. These games used to cater an overall development of children for that present day parents are paying fees to personal development centers. Indian games-indoor and outdoor-are too based on the competition in which either you win or lose, or sometimes it is a draw. Each game, besides providing fun and entertainment, also develops sensory and motor skills of the participant. Most of the traditional games are environment friendly, and they can be

played in small spaces in the vicinity. They are for all generations, and suitable for all ages.

Categories	Modern Indian Games	Traditional Indian Games
1. Space	Require more space	Require less space
2. Location	Fixed	Not fixed
3. Cost	High	Low
4. Equipments/tools, etc.	Branded products often costly	Easily accessible at low prize
5. Human Resource	More	Less
6. Gender Restriction	No	Yes
7. Mixing of age group	Partially allowed	Allowed
8. Register of games	Uniform and conform to international standard, not open to changes	Social and regional variations are available, it is open to changes
9. Goal	Played for competition and winning	Played for pleasure
10. Time	Require more time to play a game	Less time

Table 1. Modern Indian Games vs. Traditional Indian Games

3. Traditional Indian Games

India has a wonderful tradition of games which can be played by both amateur and skilled players. These games are varied from Indoor games, Outdoor games, Board games, and Property games, Language games to games based on marriage or traditional rites and rituals. North Indian games in Hindi language are different from the rest of states of the Republic of India linguistically; and they also represent festivals, social institutions, and folkloric traditions. Following are some of the traditional Indian games:

1. *Patang Baji* 'Kite flying': Children fly colorful kites with the abrasive strings. They compete with each other by cutting the competitors' kites. The register which is associated with the game is: *manja* 'abrasive strings', *girgiri* 'a tool on which stings are coiled', *sheh* 'to provide excessive stings to kite', *khech* 'to pull the strings', *khecham-tani* 'competing to capture the competitors' kite while pulling', *duggi* 'small kite', *dagga* 'big kite', *kata he-kata he* 'an expression in vocative case when you cut the strings of the competitor's kite', etc.

2. *Gilli Danda* 'bail-stick': The players hit bail with the stick. The farther your bail goes the more you score. If the competitor catches bail, or hits the stick with bail then rival is out. The register which is associated with the game is: *gicchi* 'a two inches hole in the round to place the bail', *dand* 'penalty', etc. Two or more than two players can play the game either single or in team.

3. *Kanche* 'Marbles': The glass marbles come in various colors and sizes. The players throw the marble on the ground, and he/she beats the selected marble from the opponent's choice. If you hit the target then you win otherwise you lose. The register which is associated with the game is: *gicchi* 'a two inches hole in the round to place the bail', *dand* 'penalty', etc.

4. *Satolia* 'Seven Round Stone': The players make a pile of flat stones in round shape. They try to knock them down with the rubber or plastic ball. While a player try to pile up and arrange the knocked off stones again, the opponents try to hit the player with the ball. If the ball touches you then you are out. If you succeed to arrange all the seven stones set in a sequence then you are a winner. The register which is associated with this game is: *satolia* 'I made the set', *wo maaraa* 'I hit/out.', etc.

5. *Langadi Taang* 'Hopping': The player throws a round stone in a box, and then hops on one leg in the boxes and pick up the stone without touching the boundary of the square box. The player needs to jumps over the box where the stone lies. Whoever does this for 1-10 boxes is the winner.

6. *Latto* 'Pointed Top': The players try to spin their tops as long as possible. They make a target of the opponent's top and cleave it into two. The winner is whose top spins for longer duration and cleaves the largest number of tops.

These are a few representative traditional games of India. The next section throws light on the theoretical perspective of games. The section data and discussion further discusses language games in details along with the register of other physical games.

4. Theoretical Foundations

Play, game or sport has been defined by researchers and scholars in many ways. Their criteria vary according to the purpose of their needs. But a common factor across various theoretical representations to define anything as a play, game or sport if it is an engaging and demanding activity, and it is accomplished for its own sake. The register of games is generally an uncensored activity which represents the players' spontaneous ideas created for the sake of fun. The rhyme scheme of any game is largely motivated by this function only. Freud et al. (1900) gives emphasis on the function of conflict-processing in a play. He focuses on the tension-releasing and discharging emotional elements inherently imbibed in a play which provide pleasure to a player. He states that in a play children practice the same experience repeatedly, and they gradually master it. With this self-healing process as emphasized by Freud, Children also participate in social and societal activities. Erikson (1964) believes that child develops an ability to control reality through experiments and planning during the play. He maintains a chronology based of various types of sports and how they are conducive for the development of the child.

Play is important for the cognitive development of the child. Piaget (1959) and Vygotskiĭ *et al.* (1987) consider play as a positive and leading activity respectively. Piaget states that a play helps to build emotional and intellectual equilibrium of a child where as Vygotskiĭ believes that a child is always ahead of its age and normal behavior in a play. It becomes a gateway to the Zone of Proximal Development (ZPD) for children. Classical thinkers and researchers on play postulate that it is a beneficial activity, and not only children but adults also participate and engage themselves in both rational and irrational games.

Researchers have been studying various interdependent levels-individual, group, social, societal, and historical-where human play activity occurs. Schousboe & Winther-Lindqvist (2013) make a subtle distinction between play and playfulness with reference to the actual activity of playing and player's wish to participate in the play. Any wish to participate may in turn become an unpleasant experience for the player, and many play itself can be described an unattractive experience of playfulness. In this study, however,

Dialectologia 17 *(2016), 151-166.*
ISSN: 2013-2247

the focus is on pleasant experience through voluntary engagement and participation as the data is from the children's game only. This study is supported by cultural-historical perspectives given by Schousboe and Lindqvist (2013) which purport that "the development of human beings is entangled with the developments in the surroundings they live in."

5. Data and Discussion

The data[1] of this paper for language games had been collected from many children in schools in Rajasthan and Uttar Pradesh States of the Republic of India. For ethical reasons I am unable to name the individual members or schools of the social communities. Learning a language is the first human thing that children do, and probably every normal human being masters a language in his/her initial three years of life. If we examine this learning/acquisition process we will find that there are grammatical rules and appropriateness constraint in this problem-solving/language-learning game. The registers of games that children play at later stages are assisted by the language that they acquired earlier. Moreover, cognitive games, such as riddles, ludling (secret language) or argot, and others employ language more actively than the physical and social games.

The common ludling or argot which Hindi speaking children play quite often is known as *gupt/chor bhasha* 'secret language.' The derivation of words is simple; it adds NC (nasal-consonant) and C (voiceless affricate consonant) to already existing word's initial syllable. It copies a vowel from the initial syllable, here in the example (1) it copies a long vowel *aa* from *baa* and the adds it to C structure, *e.g.*

(1) baadashah 'king' becomes baa**[-n-ch(aa)-]**dashah (usage in noun)
chidiyaa 'bird' becomes chi**[-n-ch(i)-]**diyaa 'bird' (usage in noun)

baadashah	chidiyaa	maartaa	he	'The king kills a bird.'
king	bird	kill	Auxiliary becomes	

[1] My study of children's play would not have been possible without the assistance of Ms. Gargi who made contact with girls' schools and collected data for the register of games.

ba**[-n-ch(aa)-]**dashah chi**[-n-ch(i)-]**diyaa maa**[-n-ch(aa)-]**rtaa h**[-n-ch(-)-]**e (usage in sentence)

The round brackets show the copied vowel.

There is no one satisfactory answers to the question why do children need to develop a secret language for themselves, and what pleasure do they get in it? This language game is obviously not for winning or losing but for creating a secret community among themselves.[2] But the examples clearly justify that the register of games is different from the ordinary language, and it can be coined anytime depending upon the requirement of the player. Moreover, a game may transfer from one generation to another but as human language evolves similarly the register of a language also changes. The language play by children has been of interest to psychologists and researchers. Thomas (2007) in his study points out the "carnivalesque" nature of such creativity of speech. He says that this language is probably structured by the children to undermine authorities and conventions, and employed by them out of the earshot from their parents, teachers, and other governing bodies. This creative aspect of language used by children also specify an important principle of language learning that people who communicate do not listen to each sound and word but their focus remain on the meaning instead. Moreover, the ludling does not alter the periphery of the word, and the changes occur only in the middle syllable of the word. Vygotskiĭ *et al.* (1987) states that such devices are important for appropriating meta-linguistic abilities and skills of reading and writing. He says that this ability makes the language forms opaque, and they attend to them in and for themselves. Riddle is another language-based game. It includes various literary devices, such as alliterations, rhyme-scheme, similes, metaphors, non-sense verse, etc.

Nonsense verse often includes nonsense word (NW), parody, highly rhythmic nature of words, words which are censored, tongue twisters, and dirty jokes. They cannot be defined as a game as such in terms of winning and losing but these language

[2] Since I also used this ludling in my childhood (6-7 years old), I can say that a desire to say something which only a restricted group can understand motivated me.

plays are popular among children and they generate playfulness. One such game is to find out who has fart in the group. Anyone from the group starts singing the nonsense rhyme pointing at each one after uttering each word, and the culprit is the one on which the final word *dhoos* 'an onomatopoeic word for farting' ends , *e.g.*

(2) aad**(aa)**	paad**(aa)**	kaun	paad**(aa)**	(line 1) **(aa)** rhyming
(NW)	fart	who	fart	

raam-ji	kaa	ghod**(aa)**	paad**(aa)**	(line 2) **(aa)** rhyming
God Ram	of	horse	fart	

aai	mai	dhoos	(line 3)
(NW)	(NW)	'an onomatopoeic word for farting'	

Seemingly explicit sexual overtone in general interrogative discourse marker and its reply is in fact only a rhyme scheme to generate fun for children*, e.g.*

(3) kyaa	**[huaa]**				'what happened' (a child says)
what	happened				
tere	peT	mein	bacca	**[huaa]**	'you got a baby' (another child says)
your	womb	in	child	happened	

Children also employ pun in their day-to-day language. Masti 'mischief/condom brand' is an ambiguous word, and children use it as a pun in the following example:

(4) masti		mat	kar	'Don't do mischief.' (a child says)
mischief/Condom brand		not	do	

mein	to	Kohinoor/Moods	kar raha	hoon	'I am using Moods/Kohinoor.'
I	EMP marker	Condom brands	doing	am	(another one replies)

Similarly, Kathryn Marsh (2008) recorded in her study that a group of children from Years 2 and 3 at Springfield Primary School were playing a song which was about "having sex and having a baby" without knowing the actual meaning of rhyme scheme.

Sutton-Smith's (1997) studies on play have had a significant impact on researches in play. His focus was on every conceivable perspective associated with the play. He emphasizes that play can become a delightful as well as a rude experience for the players, and they sometimes establish social order and at times violate it. Ludlings, dirty jokes, and other language games conform to this notion. Most of the language games might sound irrational and silly to the established notions of adults for games, but this initial confrontation posed by the logical/rational supporters of games can be pacified by the United Nations declaration about children's rights to play games is a human right for children. At the same time researchers and social thinkers are also concerned about the harmful impact of Modern games, such as video games, fighting games, war games, and others which are largely motivated by violent and sexual imagery. Furthermore, the addiction for modern technological games also encourages the parents and others to gear up for the traditional games.

It is not the case that elaborate rhyming scheme is only associated with language games. Multiple rhyme schemes linked with physical games explains talking-while-playing phenomenon with traditional games. In traditional games a complex elaborate musical construction juxtaposes with the game. These musical plays involve rhythmic language, movements, and text, *e.g.* a rhyme scheme associated with *Kanche* 'Marbles' goes like this:

(5)	ikkal	gaanjaa		a	(line 1)
	one	(NW)			
	dudh	pilaajaa		a	(line 2)
	milk	drink			
	tiin	ki	Tik Tik	b	(line 3)
	three	of	tic-tic		

cholam	chaaku	c	(line 4)
four	knife		
paanch ka	peindu	c	(line 5)
five of	(NW)		
chak-mak	deindu	c	(line 6)
six (NW)	(NW)		
saat ki	sutii	d	(line 7)
seven of	(NW)		
aaTh ki	uTii	d	(line 8)
eight of	(NW)		
nammak	naaraa	e	(line 9)
nine	(NW)		
das ki	dhaaraa	e	(line 10)
ten of	stream		

This game is played by two or more players. Each player throws his/her marble from a metre distance into the gicci 'an inch hole in the ground.' Whose marble is near gicci or inside it starts first. The players hit another players' marbles by making an arc with palm while putting their thumb on the ground and stretching the finger, and then gradually releases the marble. If it beats the other player's marble, the player gets the point. Any player who does this ten times first is the winner of the game. After every correct aim/hit the player sings the corresponding rhyme.

Rhythmic and formal complexity of language increases gradually from duple meter to metrical change in asymmetrical meters in their games, *e.g.* in a game mostly played by girls the rhyme goes:

(6)	akkad	bakkad	bambei	bo	(line 1)
	(NW)	(NW)	(NW)	(NW)	
	assii	nabbe	pure	so	(line 2)
	eighty	ninety	total	hundred	
	so	me	lagaa	dhaagaa	(line 3)
	hundred	in	hit	thread	
	chor	nikalkar	bhaagaa		(line 4)
	thief	come out	run		

Orff and Kodaly believe that initially children do not like study (Carder & Landis 1990). They would rather prefer to play than to learn anything. Since musical elements remain predominant in their play, Orff-Shulwerk advocates a "child-centered" play-based way of encouraging learning. Kodaly advocates that musical education is important in beginning years of formal education for children, particularly between the games of three and seven (Choksy 1981). The pedagogical implication based on these rhyming games can be observed in the tenets of educational methodology in the class-room.

6. Conclusions

The registers of children's traditional games are important not only in cultural tradition but it also paves a path for a musical educational tradition for learning. The important aspects of these register are that they are created by them to learn linguistic and meta-linguistic activities through games. Though most of the registers of games have been lost as the modern generation is forgetting them yet an effort like UNESCO's to create a Network on Traditional Sports and Games (TSG) will surely help to save this cultural heritage of India. This paper is also written in a similar direction to motivate the readers in the area of the registers of traditional language games. Extensive fieldwork

and data collection can save this tradition before falling into oblivion. However, further studies can be done on the variations of registers employed by the boys and girls in their games respectively.

References

CARDER, P. & B. LANDIS (1990) *The Eclectic curriculum in American music education: Contributions of Dalcroze, Kodály, and Orff*, Reston, VA: Music Educators National Conference.

CHOKSY, L. (1981) *The Kodaly context: Creating an environment for musical learning*, Englewood Cliffs, N.J: Prentice-Hall.

ERIKSON, E. H. (1964) *Childhood and society*, New York: Norton.

FREUD, S., J. STRACHEY, A. FREUD, A. RICHARDS & SCIENTIFIC LITERATURE CORPORATION (1900) *The standard edition of the complete psychological works of Sigmund Freud*, London: Hogarth Press.

INDIAN TRADITIONAL GAMES (n.d.) Retrieved from http://www.traditionalgames.in/ (27 August 2014).

MARSH, K. (2008) *The musical playground: Global tradition and change in children's songs and games*, Oxford: Oxford University Press.

PIAGET, J. (1959) *The language and thought of the child*, New York: Humanities Press.

SCHOUSBOE, I. & D. WINTHER-LINDQVIST (2013) *Children's play and development: Cultural-historical perspectives*, Dordrecht: Springer.

SOVIET PSYCHOLOGY: PLAY (n.d.) Retrieved from https://www.marxists.org/archive/vygotsky/works/1933/play.htm (19 July 2014).

SUTTON-SMITH, B. (1997) *The ambiguity of play*, Cambridge, Mass: Harvard University Press.

THOMAS, J. T. (2007) *Poetry's playground: The culture of contemporary American children's poetry*, Detroit: Wayne State University Press.

UNITED NATIONS GIRLS' EDUCATION INITIATIVE (n. d.) Retrieved from http://www.ungei.org/index.php (20 August 2014).

UNESCO'S NETWORK ON TRADITIONAL SPORTS AND GAMES (TSG) | UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION (n. d.) Retrieved from http://www.unesco.org/new/en/social-and-human-sciences/themes/physical-education-

and-sport/traditional-sports-and-games/unescos-network-on-traditional-sports-and-games-tsg/ (20 August 2014).

VYGOTSKĬ, L. S., R. W. RIEBER, A. S. CARTON, N. MINICK, L. S. VYGOTSKĬ & L. VYGOTSKY (1987) *The collected works of L.S. Vygotsky*, vol. 1, New York: Plenum.

Dialectologia 17 *(2016), 151-166.*
ISSN: 2013-2247

Appendix

Major traditional Indian Games

Games	Description
1. *Patang Baji* 'Kite flying' 	Children fly colorful kites with the abrasive strings. They compete with each other by cutting the competitors' kites. The register which is associated with the game is: *manja* 'abrasive strings', *girgiri* 'a tool on which stings are coiled', *sheh* 'to provide excessive stings to kite', *khech* 'to pull the strings', *khecham-tani* 'competing to capture the competitors' kite while pulling', *duggi* 'small kite', *dagga* 'big kite', *kata he-kata he* 'an expression in vocative case when you cut the strings of the competitor's kite', etc.
2. *Gilli Danda* 'bail-stick' 	The players hit bail with the stick. The farther your bail goes the more you score. If the competitor catches bail, or hits the stick with bail then rival is out. The register which is associated with the game is: *gicchi* 'a two inches hole in the round to place the bail', *dand* 'penalty', etc. Two or more than two players can play the game either single or in team.
3. *Kanche* 'Marbles' 	The glass marbles come in various colors and sizes. The players throw the marble on the ground, and he/she beats the selected marble from the opponent's choice. If you hit the target then you win otherwise you lose. The register which is associated with the game is: *gicchi* 'a two inches hole in the round to place the bail', *dand* 'penalty', etc.
4. *Satolia* 'Seven Round Stone'	The players make a pile of flat stones in round shape. They try to knock them down with the rubber or plastic ball. While a player try to pile up

	and arrange the knocked off stones again, the opponents try to hit the player with the ball. If the ball touches you then you are out. If you succeed to arrange all the seven stones set in a sequence then you are a winner. The register which is associated with this game is: satolia 'I made the set.', wo maaraa 'I hit.', etc.
5. Langadi Taang 'Hopping' 	The player throws a round stone in a box, and then hops on one leg in the boxes and pick up the stone without touching the boundary of the square box. The player needs to jumps over the box where the stone lies. Whoever does this for 1-10 boxes is the winner.
6. Latto 'Pointed Top' 	The players try to spin their tops as long as possible. They make a target of the opponent's top and cleave it into two. The winner is whose top spins for longer duration and cleaves the largest number of tops.

Dialectologia 17 *(2016), 167-199.*
ISSN: 2013-2247

Received 29 October 2014.
Accepted 10 January 2015.

REFLEXES OF THE TRANSITIVE *BE* PERFECT IN CANADA AND IN THE US: A COMPARATIVE CORPUS STUDY

Yuri YERASTOV

Fort Hays State University*

y_yerastov@fhsu.edu

Abstract

The present paper examines the geographical distribution of three reflexes of a transitive *be* perfect in North America: *done, finished, started* (e.g. *I'm done my dinner*). This paper is a corpus study of *Canadian Newsstand Complete* and *Newspaper Source Plus* — commercial databases of regional and national mass media in Canada and the United States, respectively. A total of 1217 tokens have been found in Canadian sources, and none in the US ones. Constructional tokens in Canada have been found to be proportionately distributed across provinces and municipalities. Based on these findings, it is argued that the *done my dinner* construction is a feature that sets Canadian English (as a theoretical abstraction) apart from American English. It is emphasized, though, that the construction has been attested in some dialectal pockets in the US such as Vermont. The geographical restriction of the construction to Canada and selected dialectal areas in the US is hypothesized to have arisen from a Scottish founder effect; evidence from fictional literature and migration studies is presented as support. It is further argued that the construction is a retention in North American English rather than innovation because its earliest occurrence dates back to the middle of the 19th century.

Keywords

transitive *be* perfect, geographical distribution, dialects, Canadian English, Scots, *I am done my dinner*

* 600 Park Street, Department of English, Hays, KS 67601-4099, USA.

REFLEJOS DEL TRANSITIVO PERFECTO DE *SER* EN CANADÁ Y USA:
UN ESTUDIO COMPARATIVO DE CORPUS

Resumen

Este artículo analiza la distribución geográfica de los tres reflejos del transitivo perfecto de *ser* en América del Norte: *hecho*, *acabado*, *iniciado* (por ejemplo, *he terminado mi cena*). Este trabajo es un estudio de corpus del *Canadian Newsstand Complete* y el *Newspaper Source Plus* — bases de datos comerciales de los medios de comunicación regionales y nacionales de Canadá y Estados Unidos, respectivamente. Se han encontrado 1217 registros en las fuentes canadienses, y ninguno en la de los Estados Unidos. Se han encontrado registros de construcciones en Canadá para ser distribuidos proporcionalmente entre provincias y municipios. Sobre la base de estos hallazgos, se discute que la construcción *done my dinner* es una característica que define el inglés canadiense (como abstracción teórica), diferente del inglés americano. Se hace hincapié, sin embargo, que la construcción se ha registrado en algunos puntos dialectales de los EE.UU., como Vermont. Se hipotetiza que la restricción geográfica de la construcción en Canadá y en las áreas dialectales seleccionadas en los EE.UU. puede haber surgido del efecto fundacional del escocés, según parecen indicar la literatura de ficción y los estudios de migración. Se argumenta además que esta construcción es una retención en el inglés de América del Norte más que una innovación, ya que su primera aparición se remonta a mediados del siglo XIX.

Palabras clave

transitivo perfecto de *ser*, distribución geográfica, dialectos, inglés canadiense, escocés, *I am done my dinner*

1. Introduction

One distinct feature of Canadian English that sets it apart from American English has, until recently, eluded dialectological and theoretical treatments – namely, the absence of a preposition after *done* and *finished* in exemplars such as:

(1) a. And the kids are all either finished school or finishing off university.

b. …I would come out and play ball when we were finished our work.

c. Ontario Provincial Police – when they were finished their investigation – they…

d. I'm almost finished the fifth of those ten…

e. Are you finished your question?

f. …because we'll be finished the application before…

Strathy Corpus of Canadian English

(2) a. when I don't have hockey and I'm done my homework I go there and skate

b. She doesn't receive EI when the kids are done school and grown.

Bank of Canadian English

These exemplars contrast with their Standard English equivalents *I am done with dinner* and *I am finished with my homework.* The underlying grammatical schema is not limited to *done* and *finished* only; in some Canadian dialects, it may also be extended to *started*, e.g. *I am started my homework* (Yerastov 2010, 2011, 2012). In the present paper, I will use the shorthand notation [*be done* NP] to refer collectively to the constructional schema [*be {done, finished, started}* NP].

Prior literature has reported on various reflexes of the transitive *be* perfect in Bungi English in the Canadian prairies (Gold 2007), Lumbee English in North Carolina (Wolfram 1996), and dialects spoken in Southern Atlantic states and Pennsylvania (Atwood 1953). It has also been experimentally shown that the construction [*be done* NP] is acceptable to speakers of Canadian English in Alberta, and is likely to be a special case of the transitive *be* perfect (Yerastov 2012).

The comparative corpus study, reported on in this paper, tracks the geographical distribution of the construction[1] in Canada and the US, using data from mass media databases. The results of the corpus study show that [*be done* NP] is widespread in Canadian English, and marginal in American English. The results further show that the construction is proportionately distributed across Canadian provinces and municipalities, with *done* and *finished* preferred as participles in the construction. I motivate the geographical distribution of the construction with a pattern of historical settlements, and draw connections to evidence of [*be done* NP] in historical corpora and fictional literature. In doing so, I reinforce the hypothesis of Scottish etymology of [*be done* NP] proposed in prior work (Yerastov 2010, 2011, 2012)

With a corpus focus, the present paper complements etymological conclusions drawn from experimental data (2012) but differs from the experimental work in having its main focus on the geography of the construction rather than its morphosyntax.

[1] Henceforth, I use [*be done* NP] to reductively refer to [*be {done, finished, started}* NP] unless greater exactitude is needed.

Unlike previous corpus-based work on [*be done* NP] (Yerastov 2010, 2011), the present paper is based on a substantially larger, systematically expanded sample of data.

The paper is organized as follows. In §2, I review the literature on transitive *be* perfects in English, suggesting a connection between various dialectal attestations of the transitive *be* perfect and [*be done* NP]. In §3, I discuss the sources of my data, the search procedure, and a control condition for my study. In §4, I present my findings, demonstrating the prevalence of the construction in Canada and its marginality in the US. In §5, I highlight the most important findings, and situate them in relation to some of the literary, historical, and dialectal facts; I tentatively propose a Scottish etymology for the construction. In §6, I conclude with a list of unresolved problems, and directions for future research.

2. Literature review

2.1 Periphrastic perfects and split auxiliary systems

Perfect constructions in Indo-European languages normally take the form of verbal perphrasis in which an auxiliarty – either *be* or *have* – combines with a past participle form. The emergence of periphrastic perfect constructions follows well-established grammaticalization paths (Bybee *et al.* 1994); the input to grammaticalization comes from stative constructions for *be* perfects and possessive constructions for *have* perfects. Because possession inherently requires a direct object, *have* perfects, after grammatical reanalysis, inherit that object. On the other hand, statives do not inherently require a direct object; periphrastic perfect structures that emerge from stative sources do no have a syntactic position for direct objects. Therefore *have* perfects are inherently transitive, while *be* perfects inherently intransitive. It is well known that the two periphrastic perfect schemas compete with each other for membership of past participles. In some languages, such competition results in split auxiliary systems, as for example in French or German. In others, the *have* schema takes over the perfect domain, e.g. English or Pennsylvania German (Heine & Kuteva 2005: 140-41).

Reverse developments, in which the auxiliary *be* takes over the entire perfect domain, appear to be possible as well, but they are not well – if at all – documented diachronically, being inferred from synchronic states. For example, within the Indo-European family, very few dialects/languages are known to have a fully productive tranistive *be* perfect: insular dialects of Scots – Shetlandic (Robertson & Graham 1991) and Orcadian (Flaws & Lamb 1996), Lumbee (Wolfram 1996) and Bungee (Gold 2007) English in North America, and dialects of Italian (Bentley & Eythórsson 2003: 451); but their origin remains elusive beyond speculations of cross-dialectal transmission or reanalysis of *have* as *be*. Either way, dialectally in Present Day English (PDE) we observe split auxiliary systems in which the auxiliaries *be* and *have* alternate in selecting for a limited set of transitive participles, as is seen today in Canadian English, as well as several dialects of American English.

On the synchronic level, *be~have* auxiliary splits have been approached from both syntactic and semantic perspectives or a combination thereof, but crucially they have fallen short of tackling alternations involving the transitive *be* perfect (see Yerastov (2105) for a critique of these approaches) and have little light to shed on the theoretical understanding of [*be done* NP].

2.2 Etymology of [be done NP]

Since the emergence of [*be done* NP] is not historically documented, we can use the construction's synchronic clues as a window onto its past. The construction yields resultative interpretations, as would be expected from a perfect construction in English (Yerastov 2011, 2012); for example, a speaker of Canadian English saying *I'm done the dishes* profiles the result of the event, which persists at speech and reference time. Resultative interpretation is what [*be done* NP] shares in common with its better known intransitive relative [*be done* (*with* NP)]. In fact, Yerastov (2012) notes that in PDE, the resultative meaning of the intransitive *I am done/ finished* is odd language-internally, because other dynamic verbs are incapable of such behavior. Formally similar structures such as *I am ruined, I am screwed, I am messed up* can only be understood as stative passives, in which the subject is a patient or experiencer, but never an agent. Exemplars

such as *I am done/ finished,* on the other hand, uniquely combine the two interpretations:

(3) I am finished
'I have finished' (resultative)
'My life/ career is finished' (stative passive)

Because of the similarity in the resultative semantics of the [*be done* (with NP)] and [*be done* NP] constructions, we can hypothesize that they are genetically related, and look at the history of the intransitive participles *done* and *finished* in search of insights into the common origin of the two constructions.

Is it possible that the co-existence of both resultative and stative passive interpretations in PDE, as shown in (3), represents a synchronic layering resulting from recent grammaticalization? After all, it is diachronically typical for resultatives to develop from passive statives (Bybee *et al.* 1994). However, a brief examination of historical data from the Oxford English Dictionary (OED) suggests an unlikelihood of such a scenario. OED reports the first resultative uses of the intransitive participles *done* and *finished* to have occurred rather late — at the end of the 18th century: *I was done with love for ever* (1766); *One farther favor and I am done* (1771); *The rogue is pressing and I must be done* (1776), followed by consistent attestation in the 18th century. The first resultative use of [*be finished*] is attested as late as in the 20th century: *How often have you told me that you are finished with all women!* (1939). It is unlikely that [*be* {*done, finished*}] grammaticalized from a stative passive into a resultative around the 18th century or later, because within this timeframe the *be* perfect paradigm was unproductive and likely incapable to attract new members.

In fact, such grammaticalization would not have been likely at any point in the history of mainstream British English. According to OED, the verbs *finish* and *start* are not autochtonous, Old English verbs. Derived from Old French, the verb *finish* is first attested in Middle English in 1375. While the origin of *start* is questionable, OED suggests that it is likely derived from a Scandinavian source and documents its first occurrence in 1000. A relatively late entrance of these verbs into the English lexicon

suggests they were unlikely to develop resultative uses, since at the time of their entrance the intransitive *be* perfect paradigm was already starting to lose its schematicity and was unlikely to be extended to new members, either.

More to the point, OED characterizes this construction as chielfly Irish, Scottish, US,[2] and dialectal, which suggests that it arose outside of mainstream British English, a conclusion that can be hypothetically extrapolated to the transitive [*be done* NP] because of synchronic similarity between the two constructions.

In what follows I review literature that elucidates the possibility of [*be done* NP] arising as a result of contact with Gaelic and Scots since these varieties have a *be* perfect.

2.2.1 Gaelic

Owing to its Gaelic substratum, Irish English has the *after* perfect, which combines with the auxiliary *be*. A reflex of that perfect is found in dialects of Irish English including those in eastern Canada:

(4) You're after ruinin' me
'You have ruined me'
(Filppula 1999: 90)

This *after* perfect is formally and semantically paralleled by an immediate perfect construction in Irish:

(5) Tá said tar éis teach a thógáil
Is they after house build –VN
'They are after building a house'.
'They have just built a house.'
(Hickey 2007: 149)

[2] This should probably be broadly construed as North American.

The presence of the construction [*be after* V-*ing*] in Gaelic may certainly have reinforced the usage of *be* as perfect auxiliary in the construction [*be* V-*en* NP], but a direct transformation of one into the other seems to be implausible due to a wide formal gap. Crucially, Siemund (2003) finds superstrate accounts of perfect constructions in Irish English to be more cross-linguistically plausible.

2.2.2 Scots and Irish English

A more likely candidate for the etymological source of [*be done* NP] is insular Scots. There are abundant attestations of the transitive *be* perfect in Shetlandic and Orcadian, two varieties of insular Scots. Millar (2007: 75) reports that "the most striking structural feature of Shetlandic is the use of *be* as an auxiliary verb in active perfective construction with all types of verbs". Reference works are also unanimous in recognizing the transitive *be* perfect in Shetland and Orkney. In an introduction to Shetland grammar, Robertson & Graham (1991) cite a number of instances of the transitive *be* perfect:

(6) a. Fifty voars I'm dell'd an set da tatties
Fifty spring I'm sorted and planted the potatoes
b. When A'm feenished yun A'll be dön a göd day's wark
When I'm finished that I'll be done a good day's work
c. Ye never did ony ill an' noo ye're dune me muckle guid
You never did any ill and now you're done me much good
d. I'm read my Bible.
(Robertson & Graham 1952: 11)

In a grammatical overview of Orcadian, Flaws & Lamb (1996) state that the auxiliary *be* is used instead of English *have*:

(7) a. Ah'm meed the dinner
I'm made the dinner

b. Wir biggid the stack
We're built the stack
c. Thoo'll be gotten a fair price for thee kye
You'll be gotten a fair price for your cattle
(Flaws & Lamb 1996: X)

These attestations may be independently confirmed with data from Shetlandic sources in the Scottish Corpus of Texts and Speech (SCOTS):

(8) a. heidmaster realised we were done wur [our] bit
b. A'm funn dem!
I am found them
c. A'm read dem ower
I am read them over
Scottish Corpus of Text and Speech

With regard to Shetlandic, Pavlenko (1997) hypothesizes that the transitive *be* perfect has arisen from the reanalysis of homophonous third-person singular forms of the auxiliaries *have* and *be* and subsequent paradigm leveling in favor of *be.* These processes were likely spurred by contact between Scots and Norn that dates back to the 15th century.

The productivity of the transitive *be* perfect in insular Scots and its survival to the present day leave open the possibility that there have been multiple opportunities for some of its exemplars to migrate to North America.

The restriction of the transitive *be* perfect to a set of participles *done, finished, started*, found in the present study, resembles the situation in non-insular Scots, and Irish English. One particular exemplar of the transitive *be* perfect, *I am finished it*, is so prominent in Scots that Trudgill & Hannah (1982: 88) view it as a feature of Scottish English, as opposed to Standard English exemplars such as *I am finished* and *I have finished it.* An examination of the Scottish Corpus of Texts and Speech confirms the presence of *finished* exemplars in non-insular Scottish sources:

(9) a. We're nearly finished this ain't we.
b. We are finished this ones and this ones and this ones, Mammy.
c. I'm finished something.
d. You can get ain when we're finished our tape.
e. We're just about finished the sewing.
Scottish Corpus of Texts and Speech

At the same time, *started* exemplars have been attested by Caroline Macafee as a non-standard feature of mainland Scots.[3]

(10) when I was just started school in the babies class, ...

A similar situation obtains in Irish English. Hickey (2007: 178) reports that speakers of Irish English in such geographically separated areas as Derry, Kerry, Offaly, and Monaghan counties accept the exemplar *They're finished the work now* – with a mean of 85%. While a search of the Corpus of Irish English does not yield any results, independent confirmation may be found in the Irish Internet domain (ie.), which contains all three participles *finished, started, done*.

(11) a. It is important that I am finished my morning routines by ten o'clock […].
b. I intend to continue acting classes well after I am finished the bronze award.
c. Now I am done my song, boys, but yet don't go away […]. (Waterford Songs)
d. Yes, she is done a great job […].
e. I am started the gym on Monday and plan on going 4 times a week.
Irish Internet domain (Google search)

The similarity of entrenchment patterns of [*be done* NP] in North America, Scotland, and Ireland suggests a genetic relationship.

[3] https://docs.google.com/file/d/0BzVAfXkKg9UlV2dwNERCbUwtSGc/edit

2.2.3 Reflexes in North America

Some exemplars of the transitive *be* perfect have been attested in North America (Atwood 1953, Wolfram 1996, and Gold 2007). Atwood (1953: 26-27) notices both transitive and intransitive instantiations of the *be* perfect in Southern Atlantic states – (9a) & (9b), and Pennsylvania – (9b), making a vague connection to British dialects.

(12) a. I am heard it
b. I am been thinking

More recent reports document similar occurrences in Bungi English in the Canadian prairies (Gold 2007), and in Lumbee English in Robeson county, North Carolina (Wolfram 1996).

(13) a. I am not got the horse tied upset the Hotel
b. Aw Willie, I am just slocked ['extinguished'] it the light
c. he's bin so greedy
(Gold 2007)

(14) a. If I'm got a dollar I'm got it.
b. I says, I'm Indian, I says, I'm been nothing, I says, but a Indian, I says here.
(Wolfram 1996)

Neither Gold (2007) nor Wolfram (1996) provide much descriptive evidence of the productivity of the transitive *be* perfect. Gold acknowledges that in her study of a corpus of Bungi data there are only a few unambiguous examples of the transitive *be* perfect where the auxiliary is unmistakably *be*, as in (10a) & (10b), as opposed to (10c), where *'s* is ambiguously homophonous and may represent either *be* or *have.* Likewise, Wolfram's examples are limited in productivity to the verbs *got* and *been*.

The origin of likely reflexes in North America is unsettled, as there is no direct, conclusive evidence. Wolfram (1996) seems to appeal more to language-internal resources as an explanation, while alluding, in passing, to the possibility of Scottish and

Scotch-Irish influence on Lumbee English in general. Gold (2007) points out that speakers of Bungi English in the Canadian prairies descended from employees of the Hudson Bay Company, who were systematically recruited, across several generations, in the Highlands of Scotland and the Orkney islands. She concludes that a Scottish etymology is more likely than other explanations such as vernacular universals (Chambers 2003). While neither account is conclusive in isolation, both of them point to a Scottish etymology for the transitive *be* perfect in North American English. If this conclusion is true, we should expect that the geographic distribution of the construction [*be done* NP] in North America will reflect Scottish settlement patterns.

An experimental study (Yerastov 2012) has demonstrated that the construction [*be done* NP] occurs in Canada, an area with a documented Scottish founder effect (Dollinger 2008; Benett 2003; Bumstead 1981). While most participants in the study came from Calgary and the province of Alberta, Yerastov (2012) argues that the results are generalizable to Canadian English as a whole. Other experimental studies confirm the absence of the construction in Illinois, but document its occurrence in Vermont (Yerastov 2010), another area where a Scottish founder effect has been documented (Shields 1996). Yerastov (2012, 2010) further demonstrates that the construction is not an idiom, but rather partially schematic. Thus, the subject slot is restricted to animate referents, the participle slot favors three items only, and the direct object slot tends to be marked for definiteness. However, the direct object slot is open-ended and variable. Importantly, Yerastov (2012) finds that other participles such as *made, heard, read* received marginal ratings of acceptability among his Canadian participants, a fact that suggests that there is a gradient continuum between [*be* {*done, finished, started*} NP] and other exemplars of the transitive *be* perfect, such as *I am heard deer in the bush.*

While experimental studies offer the advantage of testing the morphosyntactic limits of the construction, these studies have an inherent drawback in that they are skewed to the locale from which the participants are drawn, not to mention the observer's paradox (Labov 1972). It is understandably difficult to recruit a participant sample representative of entire North America. Even though the participant samples in Yerastov (2012, 2010) included a few consultants from areas other than Alberta, Illinois and Vermont, one needs to independently confirm the generalizability of such findings

to the entire countries of Canada and the United States. The present study attempts to do so with a corpus study of Canadian and American English. Unlike previous work, the focus of the present study is on geographical distribution; no attempt is made here to refine previous morphosyntactic and semantic analyses.

3. Methodology

3.1 Goal

The goal of the present investigation is to track the geographical distribution of the construction [*be done* NP] in the US and Canada.

3.2 Hypothesis

I hypothesize that [*be done* NP] is prevalent in Canada and marginal in the US. My hypothesis is informed by the distribution of the transitive *be* perfect across Scottish/ English dialects (Millar 2007; Pavlenko 1997; Flaws 1996), occurrences of [*be done* NP] in Canada (Yerastov 2012), and a Scottish founder effect on Canadian English (Dollinger 2008, Bennet 2003).

3.3 Sources of data

While there have been successful corpus studies of English dialects (e.g. Schneider & Montgomery 2001; McCafferty 2003; Van Herk & Walker 2005), there are still some challenges one needs to overcome in such studies. First, there are very few dialectal corpora in English, with the notable exception of the Helsinki Dialect Corpus, the Freiburg English Dialect Corpus, and dialectal annotations in the British National Corpus. Second, dialectal features are often too scarce in corpus data; a feature has to be relatively frequent, ≥ 1000 tokens per running text (Szmrecsany 2008), to afford meaningful generalizations.

To be sure, these challenges came up in my study of the construction [*be done* NP]. My searches for tokens of [*be done* NP] in well-established linguistic corpora returned marginal results, which did not correspond to the robustness of the construction, which I informally observed in Canadian and Vermont English. Thus the Corpora of Contemporary and Historical American English returned only 6 tokens, the Strathy Corpus of Canadian English – 8 tokens, and the Bank of Canadian English – 2 tokens (see Table 10 below). While these results provided valuable attestations of the construction (some of which will be discussed in § 5), they were not statistically meaningful for tracking the geographical distribution of the construction. Therefore I decided to turn to mass media databases as the source of corpus data for the present study.

My search of mass media sources was motivated by the assumption that tokens of [*be done* NP] would occur, with sufficient frequency, in regional and local newspapers. Accordingly, I chose the commercial databases *Canadian Newsstand Complete,* and *Newspaper Source Plus*. *Canadian Newsstand* is a full-text database of 355 Canadian newspapers. *Newspaper Source* is a full-text database of 729 US newspapers (along with international periodicals as well, which were excluded from search results). The comparative descriptors of the two databases are summarized in Table 1; a complete list of news media sources in these databases can be obtained from the websites of their vendors: ProQuest and EBSCOhost.

database	country	years covered	no. periodicals	no. days of coverage[4]
Newspaper Source	USA	1990-2014	729	1,525,615
Canadian Newsstand	Canada	1977-2014	355	993,775

Table 1. Comparison of databases

A particular challenge that a corpus linguist faces is the absence of information on the total number of words in news media databases. Because these databases are updated with new editions, the providers do not index the databases for word count. A corpus linguist must therefore find alternative ways of quanitifying the size of dynamic

[4] This figure is calculated by summing up the totals of days of coverage for each periodical in the database. Overlapping coverage is not subsumed but added up.

corpora, based on descriptors provided by database vendors, in order to have a meaningful yardstick against which to interpret and normalize search results.

One straightforward solution might be to measure the frequency of a linguistic feature against the number of periodicals in a database. This solution is not without problems because periodicals differ in how far they go back in time. A more accurate solution might be to count the number of days of coverage for each periodical in a database and sum up the results. Thus, one could quantify a universe of discourse covered by database material in terms of time rather than word count. Adopting this approach, we can sum up the days of coverage across all periodicals within each database (Table 1) and calculate the normalization factor for Canadian search results: 1,525,615 : 993,775 = 1.5. Of course, this course-grained approach is not without flaws either; for one, the length of a periodical varies from week to week and from publisher to publisher, and the resulting measure of corpus size is much less accurate than the traditional word count.

Yet another solution might lie in normalizing the results against another linguistic feature. Because many news media databases do not provide much-needed linguistic tools such as counting the frequency of an indvidual item, part-of-speech tagging and punctuation sensitive searching, generating frequency data for a normalizing feature may very well turn into an independent study of its own. As an automated alternative, however, we can count the frequency of database entries (articles, editorials, press releases, broadcasts) containing tokens of certain features.

But which features should we normalize against? We might try using high frequency functional words such as *the, in*, and semi-functional words such as *do* (Table 2). However, not all database engines are alike, and some will cap search results at a certain number, as does Newspaper Source, effectively rendering useless counting entries with words like *the*.

database	*the*	*in*	*do*
Newspaper Source	59,276,160	59,276,160	59,276,160
Canadian Newsstand	31,037,653	29,010,062	8,594,614

Table 2. Frequency of entries with functional words

Therefore, we should turn to somewhat less frequent words to detect variations in database size. General purpose verbs would serve that role well because they are still sufficiently frequent to generate meaningful comparisons; for example, the frequency of database entries that contain tokens of *give, take,* and *make* provide us with a meaningful comparison of database size (Table 3).

database	*give*	*take*	*make*
Newspaper Source	2,004,365	4,197,623	4,367,807
Canadian Newsstand	3,196,769	6,219,011	6,922,226

Table 3. Frequency of entries with general purpose verbs

The comparability of the two databases can be seen in a very strong positive correlation (r = 0.99) between the two sets of search results in Table 3. Further, when we calculate the normalization factors for the Canadian Newsstand, we get:

give 3,196,769 / 2,004,365 = 1.6
take 6,219,011 / 4,197,623 = 1.5
make 6,922,226 / 4,367,807 = 1.6

These results are similar to the normalization factor of 1.5 calculated on the basis of days of coverage above. Given the cross-validation of the two normalization methods proposed here, we can be assured that our comparison of Newspaper Source and Canadian Newsstand will be statistically meaningful.

3.4 Procedure

I searched the mass media databases for tokens of [*be {done, finished, started}* NP]. As a control condition I also searched for other tokens of the transitive *be* perfect with the participles *made, seen, heard, given, begun, ended, wanted, known, seen.* I used three criteria that helped me converge on this set of control verbs: 1) semantic similarity, 2) general purpose-ness, 3) prior documentation. First, I selected *made,*

begun, and *ended* because they semantically resemble *done, finished* and *started*; I avoided Romance verbs such as *terminate* and *initiate*, because formal Latinate vocabulary is unlikely to be retained in informal registers. Second, I selected *given, made, seen, known* because these are general purpose verbs, whose high frequency has been borne out by empirical studies in acquisition (2006). Considering their applicability to a large range of situations, one would expect that these verbs would be more likely to be retained during *be* perfect attrition. Third, I selected *heard, wanted,* and *seen* based on Wolfram's (1996) and Gold's (2007) attestation of these verbs in *be* perfect construction. I assumed, following Bybee (2006), that grammatical schemata are retained in memory along with lexical items; thus, there is a greater probability of these tokens resurfacing in the present-day transitive *be* perfect schema.

I entered the strings in (12) in the search engines. I excluded present-simple third-person singular forms because of potential homophony (*he's = he is* or *he has*). As a final step, I manually identified tokens with direct objects NPs and excluded duplicate search results.

(15) {am, 'm, are, 're was, were} {done, finished, started, made, seen, heard, given, begun, ended, wanted, known, seen}

The collected tokens were coded for participle choice, language user's location, and year of occurrence. Information about language user's location was intended to answer the question if the construction is restricted to any dialectal area in North America. Information about participle choice (*done, finished, started*) was intended to provide a clue about the productivity of the construction, as well as its subnationatal geographical distribution. Information about the year of occurrence was intended to clarify if the construction is an innovation or retention.

The returned results were checked for correlations with demographic data from official statistical sources, as well as another linguistic feature – the verb *give.* It was assumed that if [*be done* NP] is a general feature of either Canadian or American English, its pattern of subnational distribution would parallel that of *give*, a semantically general, dialectally unmarked verb.

4. Findings

The distribution of constructional tokens (Table 4) shows, with extreme statistical significance, that the construction is marginal in the US, and prevalent in Canada.

database	country	population[5]	# tokens	normalized # tokens
Newspaper Source	USA	313,900,000	0	0
Canadian Newsstand	Canada	34,880,000	1217	1825.5[6]

Population and tokens: N=2, r=-1, p<0.0001

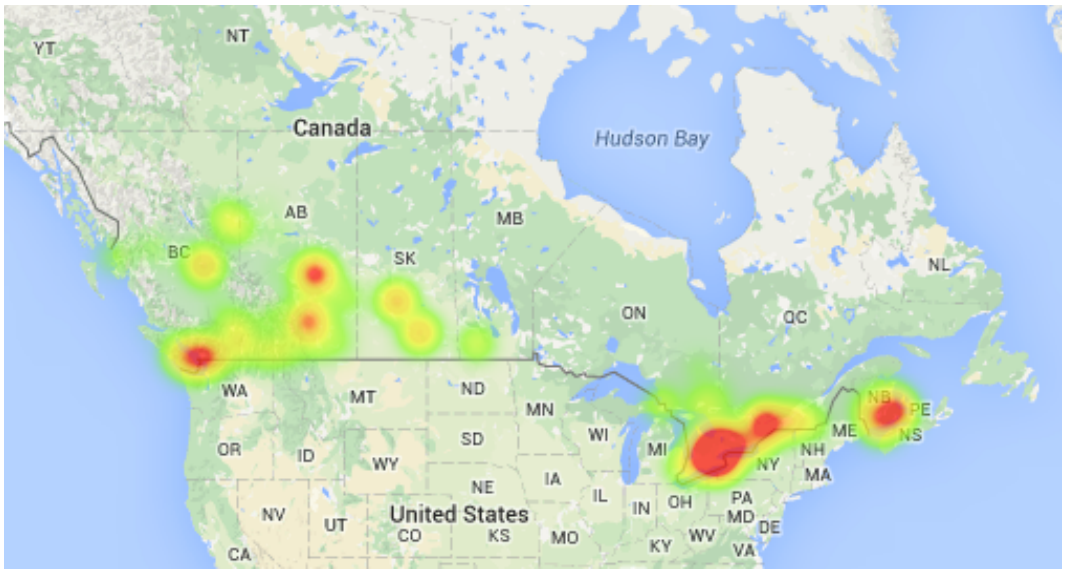
Table 4. Distribution of tokens by database

The tokens have been geocoded and are available in interactive format both as a feature map at and as a heat map at http://yerastov.fhsu.edu/maps.html. Respective screenshots are provided in Maps 1 and 2. The maps demonstrate that the construction is spread across Canada, with hotspots concentrated in densely populated areas. However, raw frequency may be misleading if not weighted for factors such as population density or frequency of other linguistic features, a topic we turn to now.

Map 1. A heat map

[5] The population figures in this article come from the US Census Bureau and Statistics Canada at the time of writing.

[6] The normalization factor of 1.5 is used to adjust the Canadian result (see 3.3 for justification).



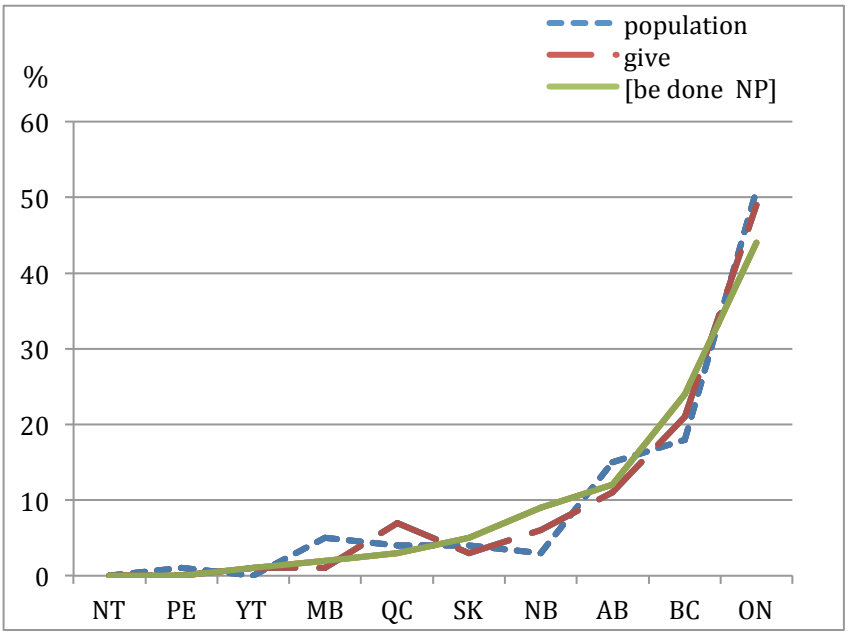
Map 2. A feature map

The results from Canadian Newsstand (Table 5 and Figure 1) show, with high statistical significance, that the construction is proportionately distributed across the Canadian provinces. When compared to the distribution of the verb *give* by province in Newsstand and the distribution of Canadian population, [*be done* NP] comes out as having an almost perfect positive correlation with those measures. Because of such close similarity in distribution with the general verb *give* (which one would presumably find in any dialect of English with comparable frequency), we can be assured that [*be done* NP] is a general feature of Canadian English. A positive correlation between *give* and provincial population validates the sampling method and confirms the internal consistency of data.

province/ territory[7]	[*be done* NP] (Newsstand)		*give* (Newsstand)		population (Statistics Canada)	
ON	531	44%	1,692,732	49%	13,505,001	51%
BC	292	24%	735,772	21%	4,622,001	18%
AB	150	12%	379,444	11%	3,873,001	15%
NB	104	9%	206,374	6%	756,000	3%
SK	66	5%	114,465	3%	1,080,000	4%
QC	35	3%	232,734	7%	1,059,001[8]	4%
MB	28	2%	43,989	1%	1,267,000	5%
YT	9	1%	29,131	1%	36,000	0%
PE	1	0%	-[9]	0%	146,000	1%
NT	1	0%	-	0%	43,000	0%
total	1217	100%	3,434,641	100%	26,387,003	100%

Table 5. Distribution of [*be done* NP] relative to *give* and population
(Canadian Newsstand and Statistics Canada)

[7] The provincial codes are: ON = Ontario, BC = British Columbia, AB = Alberta, NB = New Brunswick, SK = Saskatchewan, QC = Quebec, MB = Manitoba, YT = Yukon, PE = Prince Edward Island, NT = Northwest Territories.
[8] Anglophone population only.
[9] Data not provided by ProQuest.



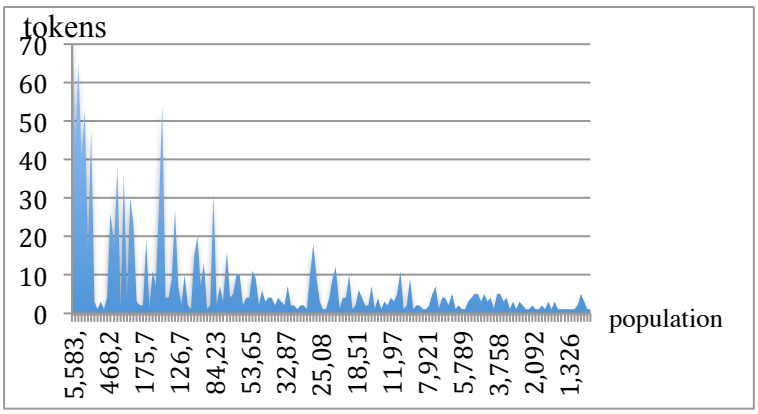
[*be done* NP] and *give*: N=10, r=0.99, p<0.0001

[*be done* NP] and population: N=10, r=0.97, p<0.0001

Figure 1. Distribution of [*be done* NP] relative to *give* and population.
(Canadian Newsstand and Statistics Canada)

Noteworthy is the high number of constructional occurrences in New Brunswick, as compared to other provinces. There are more occurrences in New Brunswick than in Manitoba or Saskatchewan, even though it has a smaller population than either of the two provinces. Similarly noteworthy is the high number of constructional occurrences in two major New Brunswick municipalities – Moncton and Saint John (4% and 2% of the entire sample, respectively). If we take into account massive Scottish immigration to the Maritime Provinces in the 18th and 19th centuries (Bumstead 1981), then these high numbers can be attributed to a lasting founder effect. The absence of any tokens from Nova Scotia, another province with a strong Scottish influence, is simply due to the irregularity of representation of Nova Scotia sources in *Canadian Newsstand* (Proquest 2013).

A positive correlation can be found between the population size of a Canadian municipality and the frequency of [*be done* NP] (Figure 2). While somewhat weaker, this correlation is similar to the one between [*be done* NP] and provincial population, thus reinforcing the internal consistency of the findings.



[*be done* NP] and municipal population: N = 162, r = 0.72, p < 0.0001

Figure 2. Distribution of [*be done* NP] by municipal population
(Canadian Newsstand and Statistics Canada)

No instances of be perfect with the participles in the control condition have been found in Newspaper Source and Canadian Newsstand. The participles that are attested are finished, done, and started, e.g.:

(16) a. When the workers are done their programs, a lot of them stay and become undocumented.

b. When you are done the game, you feel that you have gotten rid of a little stress.

c. He said if drivers find themselves behind, drivers who are done their routes early assist other trucks.

d. I am not exactly sure what this means when I first pick up the menu, and I'm not much further ahead when I am done my meal.

(17) a. When I am finished my training, I, too, will have no choice but to leave Quebec.

b. By the time she was finished university she had been prescribed all major SSRIs […].

c. In Vancouver, when people are finished work, they are finished work.

d. He's a very bright young man who was finished school and he was following his passion for skiing and the outdoor lifestyle.

(18) a. My dad passed away in 1995 when I was started my pro career in Finland and we ended up winning that year.

b. When I was started golf, Se Ri Pak won the U.S. Women's Open tournament, so this tournament is really special for me.

c. You never want to put your team down in any situation but I was started that to stick up for a teammate.

d. Western Coal was started work in the Tumbler Ridge area five years ago in an effort to tap into a surge in world coal demand […]

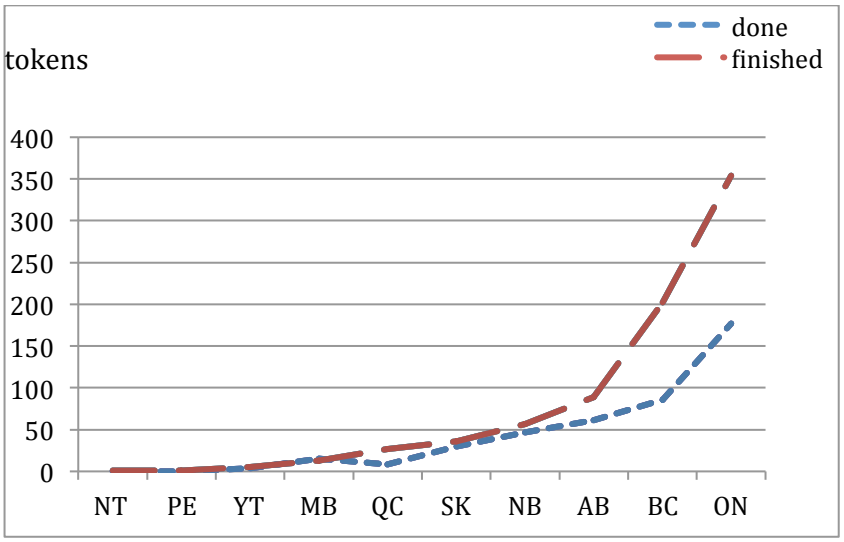
The participle *finished* dominates the Newsstand sample (64.5%), followed by *done* (35.17%) and *started* (0.25%). When assessing these findings, we should bear in mind that educated writers of English prefer to use lexically specific verbs to semantically general ones; regardless of dialect, many composition instructors have general discomfort with the form *done*. This prescptive attitude appears to hold in Canadian English and might be skewing the distribution away from *done* towards *finished* in written media, including the sample under investigation.

The internal consistency of the findings can also be revealed when we examine the distribution of the participles by province (Table 6). Excluding *started* from consideration due to its statistical marginality, we can find a perfect positive correlation between the distribution of *done* and *finished* across the provinces (Figure 3).

province/territory	done		finished		started	
ON	177	14.56%	354	29.11%	0	0.00%
BC	86	7.07%	202	16.61%	4	0.25%
AB	61	5.02%	89	7.32%	0	0.00%
NB	47	3.87%	57	4.69%	0	0.00%
SK	30	2.47%	36	2.96%	0	0.00%
QC	8	0.66%	27	2.22%	0	0.00%
MB	15	1.23%	13	1.07%	0	0.00%
YT	4	0.33%	5	0.41%	0	0.00%
PEI	0	0.00%	1	0.08%	0	0.00%
NT	0	0.00%	1	0.08%	0	0.00%
total	428	35.17%	785	64.50%	4	0.25%

Table 6. Distribution of [*be done* NP] by participle and province

(Canadian Newsstand)



done and *finished:* N=10, r=0.98, p<0.0001

Figure 2. Distribution of [*be done* NP] by participle and province.
(Canadian Newsstand)

Canadian Newsstand provides attestations of the construction from 1978 up to the present moment. A rapid increase in occurrences around the year 1998 is most likely due to the proliferation of digital technologies, which made it easier to capture news media data.

5. Discussion of findings

In what follows, I will consider whether [be done NP] is a recent innovation in North American English and adduce additional evidence to support the claim that [*be done* NP] is marginal in the US and prevalent in Canada.

5.1 Retention or innovation?

The results of the present study show that the construction has been in Canadian English since at least 1978. Has the construction arisen through innovation in the last 40 years, or had it been present in North America before?

This question may be answered with the help of the Contemporary and Historical Corpora of American English. A search of these corpora returns 6 tokens of the construction:

Token	Source	Author
Our old Doctor Bates is getting a little out of date and he'll be about ready to be put on the retired list by the time *you are done your theological course.*	The Witness (1917)	Grace Livingston Hill Lutz (1865–1947)
And it feels good to *be finished the expedition* because now I can breathe a sigh of relief that nothing really major happened to us during the expedition.	Interview with Pasquale Scaturro (2004)	Alex Chadwick
We found that Corny had not been mistaken about her influence over her family, for the next morning, before *we were done breakfast*, Mr. Chipperton came around to see us. He was full of Nassau, and had made up his mind to go with us on Tuesday.	A Jolly Fellowship (1901)	Stockton, Frank Richard
So you get into that bind. Coelho's response is, the Gore campaign *will be finished its fund-raising*.	CBS Face Nation (1999)	Jack Quinn & Ceci Connolly
I am finished 2 chapters.	No Time for Sergeants (1955)	Mac Hyman
After Scarborough, I moved to Calgary with my family -- they moved there during the boom -- and I stayed there until *I was done high school* and then I went into women's studies in Concordia University in Montreal, believe it or not.	NPR interview with singer Lullaby Baxter (2000)	

Table 7. Contemporary and Historical Corpora of American English: occurrences of [*be done* NP]

Additionally, a search of the database *Literature Online* and of the *Gutenberg Project* yields a further three instances of the construction:

Token	Source	Author
"Why, Kalman," exclaimed his sister, "*you are not half done your feast*. There are such lots of nice things yet."	*A foreigner: A tale of Saskat-chewan* (1909)	Ralph Connor (1860-1937)
I never was brought up to wait on anybody but myself. I'll go down in the yard, and play with the big yaller dog, till *they're done dinner.* That's the curiousest dog I ever did see.---I can't find out whether his tail is cut off or driv in.	*The Green Mountain Boys* (1860)	Joseph Stevens (1811-1877)
Eight! She must be done dinner by this time!	*Frank Wylde (1880)*	J. Brander Matthews (1852-1929)

Table 8. Literature Online and Gutenberg Project: occurrences of [*be done* NP]

The results of these searches, combined with the results of the present study, lead one to construct the following timeline for [*be done* NP]:

Source	Author	Year
The Green Mountain Boys	Joseph Stevens	1860
Frank Wylde	J. Brander Matthews	1880
A Jolly Fellowship	Frank Richard Stockton	1901
A Foreigner:	Ralph Connor	1909
A Tale of Saskatchewan		
The Witness	Grace Livingston Hill Lutz	1917
No Time for Sergeants	Mac Hyman	1955
Face Nation	CBS	1999
Canadian Newsstand	[multiple]	1978-2013

Table 9. Timeline of historical occurrences of [*be done* NP]

These occurrences suggest that the construction has been present in North American English for at least one and a half centuries. A retention account of [*be done* NP] is consistent with external historical facts. The construction occurs in areas such as Canada, as well as Philadelphia and Vermont, where a Scottish founder effect has been

documented (Dollinger 2008; Bumstead 1981; Bennet 2003; Leyburn 1962; Shields 1996). Unsurprisingly, New England, Pennsylvania, New Jersey, and New York are the areas from which the United Empire Loyalists moved to found the country of Canada (Dollinger 2008: 64-76). It is likely that [*be done* NP] was a feature of the dialect substratum that the United Empire Loyalists brought to Canada, which may also have been reinforced by subsequent Scottish and Irish migration to the Maritimes (Dollinger 2008: 78-88) and to Vermont (Shields 1996).

5.2 Distribution in North America

The present study shows the marginality of the construction [*be done* NP] in the US and its prevalence in Canada. This finding accords with the (scarce) data from COCA, BCE, and the Strathy Corpus. A normalized comparison of these corpora is presented in Table 10. The comparison confirms that [*be done* NP] is predominantly a feature of Canadian English.

corpus	country	corpus size (words)	# tokens	# tokens normalized
COCA	USA	450 mil.	3	3
Strathy	Canada	60 mil.	7	52.5
CBE	Canada	1.5 mil.	3	900

Table 10. Corpus search results in synchronic corpora

The quantitative marginality of the construction in corpora of US English is consistent with its scarce attestation in the US. Fruehwald & Myler (2015) attest to the presence of [*be done* NP] in Philadelphia based on native speaker intuitions. Their attestation can be independently confirmed by voluminous anecdotes on the Internet. The construction has become the subject of multiple online postings by non-linguists who discuss the "correctness" of the construction in Pennsylvania. Thus, an anonymous Internet user notices a clash between the constructions [*be done* NP] and [*be done with*

NP] in Pittsburgh vis-à-vis Philadelphia; in response, another anonymous user, who acknowledges the widespread occurrence of this construction in the Philadelphia area:

Quotation 1

[Question] How many of you use this grammatical construction -- "*I'm done my homework*." It is used by virtually everyone in the Philadelphia region [...]

[Answer] [..] Yes, I come from the Philadelphia Region. If "*I'm done my homework*" or *"I'm done the dishes"* is said around here, it sounds completely normal. […]

http://www.antimoon.com/forum/t1301-0.htm

Another dialectal area for which there is limited documentation of the construction's occurrence is Eastern New England. For example, two occurrences of [*be done* NP] were captured by the *Digital Communities Project* at Sterling College in the speech of residents native to Wolcott, Vermont (Orleans County):

(19) a. I'm all done school.

b. When he was done high school he was on his own.

Both northeastern Vermont and Phildadelphia have experienced a Scottish founder effect (Leyburn 1962; Shields 1996). For example, consider the case of Orleans County in Vermont, where several towns were settled by Scots (Shields 1996). After 1820, sixty Scottish families came from Scotland to settle in the towns of Crafsbury, Glover, and Greensboro. Pressured by the consolidation of agriculture in Scotland, they left their home, the upper Irvine Valley of eastern Ayrshire. Coming to Northeastern Vermont via New York and Montreal, they typically bought land from Yankees who were moving West (Shields 1996).

Linguistic reflections of such settlement history can be found in fictional literature. In the play *The Green Mountain Boy* (1860), the author uses [*be done* NP] in a series of Scottish linguistic features to build up the image of a local 19th century Vermonter:

(20) a. I'll go down in the yard, and play with the big yaller dog, till *they're done dinner.*

b. A sample stands *afore* you.

c. he has to *git* up about twelve o'clock every night Major, I understood you wanted to hire a chap; I s'pose a *rale* cute one.

d. I can't find out whether his tail is cut off or *driv* in.

e. I swow this grammar's awful hard stuff to *larn*.

The form *afore* 'before' occurs 1035 times in the Scottish Corpus of Texts and Speech; the form *git* 'get' – 371 times; the form *rale* 'really' – 29 times; the form *larn* 'learn' – 1 time. The form *driv* as a past participle has been attested as a feature of Scotch Irish by Montgomery (1997).

The geographical restriction of [*be done* NP] to select dialectal pockets with a Scottish founder effect suggests the likelihood of a Scottish etymology for the construction.

Further evidence supporting the Scottish etymology hypothesis may be found in the biographies of the authors (referenced in Table 9) who use [*be done* NP] in the speech of their fictional characters. These authors are, in one way or another, linked to locations with a documented Scottish founder effect: Canada, New England, Philadelphia, New York, and Georgia. Ralph Connor, born in Ontario to a family of Scottish ancestry, ministered in Alberta and Manitoba. Frank Stockton, born in Philadelphia, later in life moved to Nutley, New Jersey. Grace Lutz was a native of Wellsville, New York. Brander Matthews was literarily active in New York (although born in Louisiana); because the play *Frank Wylde* is set in New York, it is likely that [*be done* NP] reflects the language of that community. A native of Boston, Joseph Stevens Jones displayed a thorough knowledge of Scottish features in Vermont English in the play *Green Mountain Boys.* Finally, Mac Hyman was a native of Cordelle, Georgia.

The locations in which these authors were born or resided show consistent evidence of Scottish influence. Philadelphia has traditionally been a hub of Scottish migration (Leyburn 1962), where [*be done* NP] is attested today (see above); the occurrences in neighboring New York are thus unsurprising. Georgia was colonized by

Scots as early as in 1736 (Parker 1997). As discussed above, Northeastern Vermont experienced a similar Scottish founder effect – at least in the early 19th century (Shields 1996). And Canada has been influenced both by the substratum of Scottish features of the United Empire Loyalists, as well as later 19th century Scottish immigrants (Dollinger 2007).

The restricted geographical distribution of [*be done* NP] goes hand in hand with its marginal sociolinguistic status. While there is widespread acceptance of the construction by Canadian editors, the construction seems to have been stigmatized as nonstandard in the US. Particularly illustrative is a prescriptive assessment of the construction by Capital Community College in Hartford, CT, in its Guide to Grammar and Writing:

Quotation 2

"I am done my work" is completely unacceptable; I can't imagine any level of discourse where it would be tolerated. "I am done with my work" is surely acceptable in an informal setting; "I have finished my work" would be an improvement, of course.

http://grammar.ccc.commnet.edu/
grammar/grammarlogs2/grammarlogs306.htm

Such editorial taboo likely leads to [*be done* NP] being edited out of print US media – hence its non-occurrence in *Newspaper Source*.

6. Conclusion

I have shown on the basis of the present study that the construction [*be done* NP] is a century and a half old North American retention, prevalent in Canada and marginal in the US, where it is restricted to a few dialectal pockets. I have also shown that in fictional literature the construction co-occurs with other sociolinguistically marked Scottish features and that fiction writers' biographies have in common links to areas of documented Scottish influence. I have argued, on the basis of attested dialectal parallels

that the construction [*be done* NP] has a Scottish etymology, and is probably a reflex the transitive *be* perfect in insular Scots.

The results of this study confirm the generalizability of the experimental findings based on the Illinois and Alberta samples (Yerastov 2012, 2010) to Canada and the United States. While both experimental and corpus studies have their inherent methodological drawbacks, the convergence of their results contributes to our confidence that we have adequately described the distribution and properties of the construction.

References

ATWOOD, E. Bagby (1953) *A survey of verb forms in the Eastern United Stated*, Ann Arbor: University of Michigan Press.

BENNETT, Margaret (2003) *Oatmeal and the Catechism: Scottish Gaelic Settlers in Quebec*, Montreal: McGill/Queens University Press.

BENTLEY, Delia & Thórhallur EYTHÓRSSON (2003) "Auxiliary selection and the semantics of unaccusativity", *Lingua*, 114, 447-471.

BUMSTEAD, J.M. (1981) "Scottish emigration to the Maritimes 1770-1815: A new look at an old theme", *Acadiensis*, 10(2), 65-85.

BYBEE, Joan L., Revere PERKINS & William PAGLIUCA (1994) *The evolution of grammar: Tense, aspect and modality in the languages of the world*, Chicago: University of Chicago Press.

BYBEE, Joan L. (2006) "From usage to grammar: the mind's response to repetition", *Language*, 82(4), 711-733.

CHAMBERS, J. K. (2003) "Dynamic typology and vernacular universals", *Dialectology meets typology*, in Bernd Kortman (ed.), Berlin/New York: Mouton de Gruyter, 127-146.

DOLLINGER, Stefan (2008) *New dialect formation in Canada: Evidence from the English modal auxiliaries*, Amsterdam/ Philadelphia: John Benjamins.

FILPPULA, Markku (1999) *Grammar of Irish English: Language in Hibernian Style*, London & New York: Routledge.

FLAWS, Margaret & Gregor LAMB (1996) *The Orkney dictionary*, Orkney: Orkney Language and Culture Group.

FRUEHWLAD, Josef & Neil MYLER (2015) "I am done my homework - case assignment in a stative passive", *Linguistic Variation*, 15(2), 141-168.

GOLD, Elaine (2007) "Aspect in Bungi: Expanded Progressives and BE Perfects", in Milica Radišić (ed.), *Proceedings of the 2007 Annual Conference of the Canadian Linguistic Association*, Canadian Linguistic Association, 1-11 <http://homes.chass.utoronto.ca/~cla-acl/actes2007/Gold.pdf>.

HEINE, Bernd & Tania KUTEVA (2005) *Language contact and grammatical change*, Cambridge: Cambridge University Press.

HICKEY, Raymond (2007) *Irish English: History and present day forms*, Cambridge: Cambridge University Press.

LABOV, William (1972) *Sociolingustic Patterns*, Philadelphia: University of Pennsylvania.

LANGACKER, Ronald W. (2000) "A dynamic usage-based model", in Michael Barlow & Suzanne Kemmer (ed.), *Usage-based models of language*, Stanford: CSLI, 1-63.

LEYBURN, James G. (1962) *The Scotch-Irish: A social history*, Chapel Hill: The University of North Carolina Press.

MCCAFFERTY, Kevin (2003) "The Northern subject rule in Ulster: How Scots, how English?", *Language Variation and Change*, 15, 105-139.

MILLAR, Robert McColl (2007) *Northern and Insular Scots*, Edinburgh: Edinburgh University Press.

PARKER, A.-W. (1997) *Scottish highlanders in colonial Georgia*, The University of Georgia Press: Athens, GA.

PAVLENKO, A. (1997) "The origin of the be-perfect with transitives in the Shetland dialect", *Scottish Language*, 16, 88-96.

PROQUEST 2013, Canadian Newsstand™. <Retrieved 03/21/2013 from http://www.proquest.com/enUS/catalogs/databases/detail/canadian_newsstand.shtml>.

ROBERTSON, T. A. & John J. Graham (1991) *Grammar and usage of the Shetland dialect*, Lerwick: The Shetland Times Ltd.

SCHNEIDER, Edgar W. & Michael MONTOGOMERY (2001) "An electronic corpus of Southern U.S. ante-bellum overseers' letters", *American Speech*, 76(4), 388-410.

SIEMUND, Peter (2003) "Substrate, superstrate and universals: Perfect constructions in Irish English", *Dialectology meets typology*, Bernd Kortman (ed.), Berlin/New York: Mouton de Gruyter, 401-434.

SIMPSON, J. A., E. S. C. WEINER & OXFORD UNIVERSITY PRESS (1989) *The Oxford English Dictionary*, Oxford: Clarendon Press.

SHIELDS, Bruce P. (1996) "Scots among the Yankees: The settlement of Craftsbury East hill", *Vermont History*, 64(3), 174-183.

SZMRECSANYI, Benedikt (2008) "Corpus-based dialectometry: aggregate morphosyntactic variability in British English dialects", *International Journal of Humanities and Arts Computing*, 2(1-2), 279-296.

TRUDGILL, Peter & Jean HANNAH (1982) *International English: A guide to varieties of Standard English*, London: Edward Arnold.

VAN HERK, Gerard & James A. WALKER (2005) "Regional variation and early African American correspondence", *Language Variation and Change*, 17, 113-131.

WOLFRAM, Walt (1996) "Delineation of and description in dialectology: The case of perfective I'm in Lumbee English", *American Speech*, 71.1, 5-26.

YERASTOV, Yuri (2010) '*I am done dinner': When synchrony meets diachrony*, PhD diss. University of Calgary.

YERASTOV, Yuri (2011) "Grammaticalization and lexicalization of the resultative marker done in Scots-influenced varieties of North American English", in Patricia Sutcliffe (ed.), L*ACUS XXXVI: Mechanisms of Linguistic Behavior*, Houston, TX: LACUS, 331-346.

YERASTOV, Yuri (2012) "Transitive be perfect in Canadian English: An experimental study", *Journal of Canadian Linguistics*, 57(3), 1001-1031.

YERASTOV, Yuri (2015) "A construction grammar analysis of the transitive be perfect in present-day Canadian English", *English Language and Linguistics*, 19(1), 157-178.

Dialectologia 17 *(2016), 201-226.*
ISSN: 2013-2247

REVIEW ON THE 8th CONGRESS OF THE INTERNATIONAL SOCIETY FOR DIALECTOLOGY AND GEOLINGUISTICS

Eastern Mediterranean University, North Cyprus, 14-18 September, 2015

Shunsuke OGAWA[1] & Javanshir SHIBLIYEV[2]

Prefectural University of Hiroshima[1] / Eastern Mediterranean University[2]

bach@pu-hiroshima.ac.jp / javanshir.shibliyev@emu.edu.tr

Introduction

The Eastern Mediterranean University hosted the 8th edition of the International congress which is organized every three years by the International Society for Dialectology and Geolinguistics (SIDG). Most recent editions had been held at Maribor (2009) and Vienna (2012)[1]. Ahmet Pehlivan and Vugar Sultanzade perfectly leaded their organizing committee. Thanks to their dedicated efforts, participants of this congress could spend meaningful and fruitful time comfortably.

Cratos Premium Hotel, a luxurious five-star holiday resort hotel, was chosen as the main venue. Blue Hall and three beautiful classrooms of Eastern Mediterranean University were used as the special venue on September 15. The information on the website of the University informs that 200,000 students from 106 different countries are enrolled in this university. In fact, we saw students of different nationalities in the campus. It seemed to be that North Cyprus is a Suitable location for the international congress for language and linguistics.

[1] Reviews on these congresses were published on *Dialectologia* 4 & 10. The first one was written by Ernestina CARRILHO & Sandra PEREIRA on 2010, and the second one was written by Borja ARIZTIMUÑO & Lorea UNAMUNO in 2013. So this is the third paper which reviews congress that was organized by the International Society for Dialectology and Geolinguistics.

Eighty-two researchers from seventeen different countries attended this congress, and 76 lectures were given[2]. The official languages of the conference were English and Turkish. During the congress, two sessions were carried out simultaneously. One was an English session, and another was a Turkish session, but in the afternoon of the second day, three sessions were done concurrently.

1. Papers presented in English

This section reviews only English sessions.[3]

The topics of the congress covered all aspects of dialectology and geolinguistics. Name of the topics which had been posted on the web site for this congress were as follows.

- New methodological and technical approaches to variety linguistics & language geography.
- Interdisciplinary dialectology.
- Dialect synthesis
- Quantitative dialectology
- Cognitive linguistics and dialectology
- Dictionaries, atlases
- Dialect corpora, standards and norms, infrastructure
- Online dialectology
- Perceptual dialectology
- Dialect dynamics between standardization and dialectalization
- Minority languages, interference phenomena
- Dialect and culture as well as dialect and cultural studies
- Dialect and youth language
- Dialect and media (linguistics)

[2] This information was given by a secretary of SIDG.

[3] Because I could not attend a few lectures in English session, I am not able to write review about them.

- Dialect translation
- Science intermediation
- Research reports

1.1 Keynote lectures

Four excellent researchers were invited as the keynote speakers, but regrettably a certain scholar could not attend. The first invited speaker was Eva Agnes Csato who delivered the talk which entitled *Five Dimensions of Dialectological Distance*. She claimed that five different dimensions are useful to measure dialectological distance. Nowadays, many researchers make efforts to measure the dialectological distance objectively. Her new suggestion will/should be examined by other scholars for further sophisticated methodology of Dialectology.

The second invited speaker was Michael Schulte, on *Scandinavian dialects in the first millennium - Where do we stand today and what do we know*?. As the title of his lecture was easy to understand, audience who had little information and knowledge about Scandinavian dialects could gain the newest insight about them.

The last invited speaker was Nurettin Demir. He carried out the talk entitled *Küreselleşme, Standart Türkçe ve Türkçenin Ağızları* and introduced their complex language situations which include language contact, language change and variation between elderly people and younger generation, the relationship between immigration and language, standardization and so on[4].

[4] Originally, he was going to do his presentation in Turkish, however he used English instead. Thanks to his compassion, many participants who could not understand Turkish got to understand his interesting and meaningful lecture.

1.2 Lectures[5]

Lectures were classified into four topics as follows:

- 2.1 Geolinguistics
- 2.2 Sociolinguistics
- 2.3 Descriptive linguistics, Documentation, Dictionary and Corpus
- 2.4 Dialect education

1.2.1 Geolinguistics

Takuichiro Onishi (*Starting Place and Diffusing Area of Language Changes*) pointed out 2 facts by comparison of interval data of dialectal distribution: a) The places where new language changes occur are not necessarily urban or central. b) The diffusing patterns of dialectal distribution are not radial like ripples, but seem to occupy the areas inhabited by human communities.

Vilja Oja (*Some Perspectives of Mapping the Semantic Variation*) focused on areal relations of word meanings in cognate and contact languages. As she mentioned in her abstract, the method of mapping semantic variation of a word was not often used in traditional geolinguistics.

Shunsuke Ogawa (*Who Is the "Old" Believer in Japan? — Historical Sociolinguistic and Geolinguistic Aanalysis*) interpreted his geolinguistic data gathered by postal survey for Japanese Catholic Church. He claimed that historical and social events in each region had changed the meaning of *Kyushinzja* (Old Believer).

Tsunao Ogino (*Automatic Isogloss Drawing*) developed a computer program to draw the isogloss and applied it to the dialect data published by some Japanese researchers. The results showed that the program generated valid and satisfying isoglosses to the data. He introduced the idea of automatic isogloss drawing and some results of isoglosses drawn by the program.

---

[5] Introductions of each lecture are frequently quoted from *Abstracts of VIIIth congress of the International Society for Dialectology and Geolinguistics* edited and published by Organizing Committee for VIIIth Congress of the International Society for Dialectology and Geolinguistics in 2015.

Sheila Embleton, Dorin Uritescu & Eric S. Wheeler (*Exploring Linguistic vs Geographic Distance Quantitatively*) made a quantitative case study using data from a region of Romania where the geography provides more than one way of measuring distance: travel distance, and travel time between locations are not the same as the direct distance. Their geographic distances are made, in part, using Google to get travel distance and travel time. The quantitative measures are then analyzed using the R Statistics package, to provide a statistically sophisticated answer, with relative ease.

Dorin Uritescu, Sheila Embleton & Eric S. Wheeler (*Analyzing Dialect Variation with Metadata: The New Format of the Romanian Online Dialect Atlas – Crişana*) introduced their new approach to analyze dialect variation according to different morphological and syntactic aspects of *Romanian Online Dialect Atlas* (RODA[6]).

Jožica Škofic (*Interactive Slovene Linguistic Atlas and on-line Dictionaries of Slovene Language*) demonstrated on-line lexicographical and cartographical presentation of Slovene dialectal words. By this progressive approach, anyone who can use the internet is able to hear and see the recorded dialectal material — with access to digitized archive material, audio and video recordings or other on-line links to information on local dialect studies and places.

The object of a lecture by Danguolė Mikulėnienė (*The Change of Differential Dialectal Features in The New Lithuanian Dialectal Derivatives*) was the change in distinctive features of Lithuanian dialects in the beginning of the 21st century. The traditional Lithuanian dialect classification is based on the phonetic and phonological features of dialects and they are changed by other features (for example, the different pronunciation of the long unstressed vowel o).

From her ongoing project, Daiva Aliūkaitė (*The Geolinguistic Competence of Young Lithuanians*) introduced the geolinguistic competence of young Lithuanians. She explained that geolinguistic competence is understood as knowledge and skills in recognizing the variability of language and relating it to the geographical areas, social areas and cultural areas.

---

[6] *Romanian Online Dialect Atlas* which was edited and published by Sheila EMBLETON, Dorin URITESCU & Eric S. WHEELER in 2007 is able to be accessed on web. http://pi.library.yorku.ca/dspace/handle/10315/2803.

Asta Leskauskaitė (*The Southern Aukštaitians at the Beginning of the 21st Century: Dialectal Continuity and Language Attitudes*) reported about a project named *Modern Geolinguistic Research in Lithuania: the Optimization on Network Points and the Interactive Dissemination of information*. The data of the sociolinguistic questionnaire and audio recordings of 370 representatives of the Southern Aukštaitians from almost 80 localities have been evaluated and summarized.

Violeta Meiliūnaitė (*Life of Dialects: Living Place Vitality vs Dialect Vitality*) gave an overview of vitality of Lithuanian dialects in the beginning of 21st century and defined the most important tendencies of Lithuanian dialectal map development. This lecture is an outcome of the global grant project mentioned in review for Asta Leskauskaitė's lecture. The vitality of dialects is dealt with by Fumio Inoue & Yasushi Hanzawa's lecture, too.

Markus-Narvila Liene (*Language Contacts in the Region of South-Western Kurzeme*) dealt with language contact in South-Western Kurzeme. This region, situated on the shore of the Baltic sea near the border of Lithuania, is a territory of distinct cultural and regional identity, where language contacts and borrowed vocabulary have always played an important role. Lithuanian borrowings, German borrowings, Slavic borrowings and Livonian borrowings were introduced in detail.

The study of Christina Schrödl & Barbara Piringer (*Language Contact in the Dictionary of Bavarian Dialects in Austria – An Analysis of Volumes 1 to 5*) dealt with the constitution of the Australian Bavarian vocabulary in the Dictionary of Bavarian Dialects in Austria (WBÖ). They said that a specific feature of the WBÖ is the documentation of borrowings from Australian Bavarian dialects into neighbor languages like Hungarian or Slovene. Therefore, it is possible to analyze borrowings from the Bavarian dialects into the neighbor language as well as from donor languages and donor dialects into the Austrian Bavarian dialects[7].

Keiichi Takamaru & Akemi Yamashita (*Study on Phonetic Variation of Modern Japanese Loanwords*) mentioned that Japanese loanwords are originally European words and are written in *Katakana*, which is Japanese phonetic letter. They confirmed

[7] Web address of Institut für Österreichische Dialekt- und Namenlexika, Geschichte is as follow. http://www.oeaw.ac.at/icltt/dinamlex-archiv/Geschichte.html

3 hypotheses by speech data of Recording Survey of Japanese Dialectal Sounds. a) *Katakana* of loanwords have diachronic variations, e.g., /i N ki/ and /i N ku/ (ink) which means that pronunciation of loanwords may have generational variations. b) Pronunciation of original language may influence pronunciation of a loanword. c) Japanese has dialectal differences within phonological system. It may influence pronunciation of a loanword.

Shigeko Sugiura (*Dialectal Variation Regarding the Sentence-Final Particle ga in Japanese*) reported on sentence-final particle *ga*. She concluded that *ga* originally appeared following conjectural forms, and in some dialects it came to be used following declarative forms in the reminding function. Its function was reanalyzed, and came to be used in the informing function, and that in dominant areas, there arose other forms which express informing.

1.2.2 Sociolinguistics

The intermediate varieties usually are summarized under the term '*tussentaal*' (literally 'in-between-language'), an academic term for the Dutch 'everyday language'; '*tussentaal*' is used by the larger part of the Flemish population in informal situation. Jacques Van Keymeulen (*Implicational scales in the Flemish 'everyday language' ('tussentaal')*) insisted that the case that the characteristics of the Dutch 'everyday speech', used by Flemings in informal situation, might be grouped together in different 'levels' with regard to their distance to the standard language. He commented on the difference between the levels and investigated the implicational structure in them, on the basis of a corpus of spoken language. He also said that implicational scales might indeed reveal structure in the variability of the everyday language continuum in Flanders.

Danila Zuljan Kumar (*Identity Process Changes in the Slovenian and Friulian Linguistic Communities in Friuli–Venezia Giulia, Italy*) demonstrated how the attitude of the Slovenians and Friulians in the Italian Autonomous Region towards the use of their local and regional languages in public life has been gradually changing, under the

influence of attitude changes to minority, regional and local language on the part of the national and European public authorities in recent years.

As she mentioned in her abstract, living in a foreign linguistic environment, it is difficult to avoid the impact of other languages. Regina Kvasyte (*Specificity of the Spoken Language of Latvian Lithuanians: the Use of Verbs*) reported on Spoken Language of Latvian Lithuanians. She said that in the case of Lithuanian as a native language could be impacted by the Latvian language as a dominating language of state residents. She also introduced some certain specificity such as choosing verbs and using semantically incorrect verbs in a sentence in the spoken Lithuanian language of Latvian Lithuanians.

Anna Stafecka (*Main Tendencies in the Dynamic of Latvian Dialects in the 21st Century*) introduced a research project named *Latvian Dialects in the 21st Century: a Sociolinguistic Aspect* and explained some results of the project. The main objective of the project is to gain an insight into the current situation of Modern Latvian sub-dialects, and to see how their maintenance is influenced by such factors as the ethnicity, education, occupation, and age of the speakers, as well as the local cultural environment and the distance from the nearest urban centers, churches, and cultural or educational institutions.

On past tense and past participle forms of English, Heinrich Ramisch (*Past Tense and Past Participle Forms from a Variationist/Phonetic Point of View*) discussed the relationship between the written verb-forms and their actual pronunciation in the spoken language. The result of the study was carried out with a group of 20 American undergraduate students.

Keiko Hirano (*Grammatical Variation in a Dialect Contact Situation: Accommodation of Verbs of Obligation*) reported on dialect contact and linguistic accommodation in the use of verbs expressing obligation (such as 'must', 'have got to', 'have to' and 'got to') among native speakers of English who are living in Japan, using a social network approach.

Fumio Inoue (*Contradictory Tendencies of Real and Apparent Time Changes — Late Adoption of Politeness Behavior in Okazaki Honorifics Survey*) reported on the results of the surveys which the NINJAL implemented repeatedly over more than half a

century in a city in Central Japan. Honorific expressions and politeness behaviors for asking and introducing matters were investigated through interviews, and whole utterances were transcribed. The results of the analysis revealed that discourse behavior and honorifics are acquired late in life.

Fumio Inoue & Yasushi Hanzawa (*Observation of Linguistic Change in Progress Through Real Time Comparison of Glottogram Data*) compared the results of two Glottogram surveys. Glottogram is a technique developed in Japan, showing geographical location on one axis and age of consultants on the other axis. They proved that Japanese dialects are still strong and have possibilities of survival against the commonsense of Japanese people.

Klaus Geyer (*Semi-Communication Within the Scandinavian Dialect Continuum as a Challenge for Audio-visual Translation: the Case of the Danish-Swedish Crime Series Broen – Bron (The Bridge)*) discussed translation strategies from the Danish-Swedish Crime Series named *Broen – Bron (The Bridge)*. To consider the strategy, he compared translated edition (Swedish, Danish, English and German (dubbed) of the TV drama.

1.2.3 Descriptive linguistics, Documentation, Dictionary and Corpus

Alexander Mankov (*The Dialect of Gammalsvenskby: an Outline of Verb Morphology*) showed us very interesting linguistic data (verb morphology) used in Gammalsvenskby which is a village located in Ukraine. The dialect is the only surviving Scandinavian dialect in the territory of the former Soviet Union. His lecture attracted attention because participants of this congress did not know about the dialect.

Maria Pupynina (*Dialectal Variation in Chukchi-Kamchatkan Language Family*) introduced dialectal variation in *Chukchi-Kamchatkan* (Chukotko- Kamchatkan) — a small language family whose languages are spoken in the Far East of Russian Federation. The family includes language: *Chukchi*, *Kerek*, *Koryak*, *Alutor* and *Itelmen*. She emphasized that there is a strong need for a revision of dialectal variation in Chukchi-Kamchatkan family and an urgent field investigation of Chukot and Koryak

language groups (including perceptual dialectology) because all these languages are seriously endangered. Her lecture attracted attention as well.

Tokunoshima is an island in the Ryukyu Archipelago. Chitsuko Fukushima (*Reorganization of Verbal Conjugation System in Japanese Dialects: a Case Study in Tokunoshima Dialects*) reported reorganization of verbal conjugation system in Tokunoshima Dialects which are now facing the risk of extinction. She concluded that originally there are 3 conjugation systems but a conjugation system has been disappearing in many dialects recently.

Motoei Sawaki, Yumi Nakajima & Chitsuko Fukushima (*Practical and Academic Assets of Multimedia Dialect Dictionary Based on Sentence-based Corpus*) have been involved in making of dialect dictionary based on sentence-based corpus. They made new multimedia dictionary which has two different kinds of assets. One asset of the dictionary is that it is useful for the users (above all younger generation) to get the practical command of the dialect. Another is academically valuable asset.

Maria-Pilar Perea (*Putting Together the Sources of a 20th-Century Dialect Catalan Dictionary*) introduced a project entitled "Computerization, dialectal sources, lexicographical influences, mapping and sound of the *Diccionari català-valencià-balear* (DCVB2.0+). The main objective of the project is to get interactive searches that helps the development of more comprehensive lexical studies, based primarily on the collection of documentary sources that became the written aspect of the work.

Gotzon Aurrekoetxea (*The Determinants of Language Variation*) proposed two statistical procedures to determine the following: on one hand, he showed which linguistic features have caused the greatest variation in the different clusters or dialect areas; and on the other hand, he showed the localities which have the largest number of linguistic features of the analyzed linguistic area.

Miho Saito (*Cause or Trigger? -The different ways of categorization in Japanese variations*) reported difference in categorizing conditional and casual relationship between Standard Japanese and Setouchi-dialect, one of the Ryukyuan dialects spoken in southern area of Amami islands in Kagoshima prefecture.

1.2.4 Dialect education

Genovaitė Kačiuškienė (*Unique Series of Publications "Lithuanian Dialects for School": Virtual Information in Web Portals*) introduced a series of publications named "*Lithuanian Dialects for School*". The series consists of study books with compact discs for teachers and students, which help students to expand their horizons, enrich the vocabulary and improve knowledge about self-consciousness, the culture and history of their native land, as well as reveal the beauty inherent in spiritual creative works by representative of various dialects. Such project will become more important among any regions or countries in the age of globalization.

1.3 Conclusion

As above description of chapter 2, I introduced each lecture individually. I will briefly comment on all lectures.[8]

Geolinguistics was a topic discussed mostly in this international congress. There was same tendency in 5th & 6th & 7th congress of the International Society for Dialectology and Geolinguistics (SIDG)[9]. On the other hand, it seems to be that the interest for Sociolinguistics is increasing. The keywords of Sociolinguistics in this congress were younger generation, identity, language contact, grammar and so on.

Many researchers from Baltic countries and Japan participated in this congress. Thanks to them, we familiarized ourselves with Lithuanian, Latvian and Japanese. However, I hope that more and more researchers from various regions and countries will take part in the next congress. Next congress of SIDG is scheduled to be held at Vilnius, Capital of Lithuania, in the end of July, 2018.[10]

Lastly, I want to express my sincere thanks to Ahmet Pehlivan and Vugar Sultanzade (Head of the organizing Committee).

---

[8] Proceedings of this congress will be published in the near future by organizing Committee.
[9] I attended these congresses.
[10] http://www.geo-linguistics.org/conferences.html

References

ARIZTIMUÑO, Borja & UNAMUNO, Lorea (2013) "Review on the VIIth Congress of the International Society for Dialectology and Geolinguistics (SIDG) & WBOE^{100}", *Dialectologia*, 10, 135-146. <http://www.publicacions.ub.edu/revistes/dialectologia10/reviews.asp>

CARRILHO, Ernestina & PEREIRA, Sandra (2010) "Review on the VI^{tth} Congress of Dialectology and Geolinguistics", *Dialectologia*, 4, 111-120. <http://www.publicacions.ub.edu/revistes/dialectologia4/reviews.asp>.

ORGANIZING COMMITTEE FOR VIIITH CONGRESS OF THE INTERNATIONAL SOCIETY FOR DIALECTOLOGY AND GEOLINGUISTICS (eds.) (2015) *Abstracts of* VIIIth *Congress of the International Society for Dialectology and Geolinguistics*, Famagusta: Eastern Mediterranean University.

2. Papers presented in Turkish

The rationale behind the choice of the venue of the 8th Congress of the Society was quite evident. Metaphorically speaking, Cyprus is a paradise for linguists, especially for those who study dialects synchronically and diachronically. Throughout its history, Cyprus has been invaded by many ethnic groups whose languages left their language relics in the Turkish Cypriot variety being spoken in North Cyprus today. Moreover, due to its being an island Turkish Cypriots have lived in isolation caused by physical barriers and, therefore, the Turkish Cypriot dialect has managed to preserve many linguistic properties that are inexistent in the Standard Turkish or in the other Turkish regional varieties. It is also fact that language situation with various forms of contact in Cyprus is unique. In short, when walking in the streets of Cyprus, one can easily observe the diversity of linguistic soundscape and landscape. This must have also contributed to the huge interest in the congress: approximately eighty researchers from about twenty different countries presented their papers to the enthusiastic audience. The working languages of the conference were both English and Turkish.

The scope of the conference was quite diverse. The participants addressed various aspects of dialectology and geolinguistics. Among the aspects considered were:

innovative methods and approaches to dialect analysis, dialect-culture-ethnicity triangulation, history of specific dialects, dialect contact, etc. Some of these aspects were more frequently addressed in papers presented in Turkish.

2.1 Keynote speakers

Among the keynote speakers there were such outstanding researchers in the field as Prof. Dr. Lars Johanson, Prof. Dr. Eva Agnes Csato, Prof. Dr. Michael Schulte and Prof. Dr. Nurettin Demir. The topic of the paper to be presented by Prof. Dr. Lars Johanson and Prof. Dr. Eva Agnes Csato was about the dimensions of dialectological distance. As Prof. Dr. Lars Johanson had not been to attend the congress, the coauthor Prof. Dr. Eva Agnes Csato presented the paper which focused on five dimensions of dialectological distance. Her suggestions triggered huge interest in the audience. The second keynote speaker, Prof. Dr. Michael Schulte, laid his view of the Scandinavian dialects in the first millennium and dealt with the state-of-the-art in the research on these dialects. Although the majority of the audience was not sufficiently knowledgeable about Scandinavian dialects, the paper attracted great interest. The third paper was intended to be presented in Turkish. However, the topic instigated much interest, and it was suggested that the paper be presented in English so that multilingual audience could have a chance to listen. Thus, Prof. Dr. Nurettin Demir addressed both Standard Turkish and dialects of Turkish vis-a-vis globalization. Emphasis was placed on language contact, language change and the role of age vis-à-vis change and immigration. It was interesting to consider dialects from such perspectives.

2.2 Lectures

Lectures delivered in Turkish can be categorized under certain titles.

2.2.1 Onomastics of literary works

Aynure Eliyeva *("Kitab-i Dede Korkut" Destanı İle Bağlanan Karabağ'in Bazi Toponimlerinin Etimolojisi Uzerine*) analyzed the etymology of place names used in the Turkic epic poem "Kitab-ı Dede Korkut" and compared them with similar onomastic units which can be seen as place names in Garabagh region of Azerbaijan and in other territories inhabited by various ethnic groups. The driving force for the writer to focus on this topic was the importance of place names as an indicator of the history of certain geography. The author found addressing this topic more meaningful now when Armenian occupants purposefully change place names in the region to distort its true history.

The presenter underscored the relationship between the toponyms concerning the region and the toponyms used in the Turkic epos of "Kitab-ı Dede Korkut". The writer found it quite natural since both toponyms of the region and the words used in the epic poem derive from the same origin. In order to support her view, she focused on such toponyms in the given area as Paşabēyli, Sarıcoban, Umudlu, Eyvazhanbeyli, Eliağalı, Hacımemmedli, İmamgulubeyli. It should also be mentioned that similar place names in other parts of Azerbaijan and in some other Turkic territories were subjected to analysis. The writer concluded that the etymological analysis of toponyms in Garabagh points to the fact that the traces of various Turkic ethnic groups can be observed there.

Elvan Ceferov *("Dede Korkut" Destani ile Doğu Anadolu'nun Erzurum, Kars, Iğdir ve Ardahan Dialektlerinin Ortak Leksikoloji Ozellikleri*) focused on the lexical properties of the language of the Turkic epos "Kitab-ı Dede Korkut" and attempted to compare its lexical properties with those used in Eastern Anatolian dialects of the Turkish language. The presenter underscored the view that the language used in the epos of "Kitab-ı Dede Korkut" cannot be associated with a limited territory; rather its language structure reflects the linguistic features of the old Azerbaijani and Anatolian Turkish. In other words, most lexical units used in "Kitab-ı Dede Korkut" can be encountered in the eastern Anatolian dialects of Turkish. The researcher also dealt with words and

suffixes in “Kitab-ı Dede Korkut” that reflect linguistic features of the eastern Anatolian dialects. It was also mentioned that, when comparing the language of Dede Korkut with Western Anatolian dialects and old Anatolian dialects, it could be seen that it has its own specific vocabulary stock and stylistic features. According to the presenter, the epos, whose language reflects the first period of the Western Anatolian Turkish and Old Anatolian Turkish, is of great importance for the Azerbaijani language and culture.

It was also found out that many lexical features of “Kitab-ı Dede Korkut” can be encountered in such Eastern Anatolian cities as Erzurum, Kars, Igdir and Ardahan. The writer provided examples from these dialects to support his views. Ceferov explained this with the strategic position of Kars which served as a transition point for immigration from Caucuses to Anatolia. The researcher believed that the wide diversity of dialects in this geography owes much to relocations of various ethnic groups. The presenter reiterated his view that the epos belongs to all Turks and especially to the regions described here.

Zemfira Abbasova *(Elcin Efendiyev'in Yazılarında Kullandığı Onomastik Adlar Ve Onların Gruplandırılması*) analyzed the onomastic units used by Azerbaijani writer E. Efendiyev in his literary works. The researcher stressed the importance of onomastic approach as etymological and structural analysis which may provide valuable data about the origins of an ethnic group, its history and social life and its spiritual world. The author underscored the role of literary works as a main source to get insights into the field of onomastics of language. The main emphasis in the analysis was placed on such categories as anthroponyms, toponyms and ktematonyms which frequented in Efendiyev’s literary works. It was found out that Efendiyev uses both real onomastic units and fictitious/fantastic ones. Analysis of anthroponomy in terms of their origins showed that, along with native units, units originated from Arabic, Persian, Russian and European origins are also widely used. Anthroponyms were also subjected to morphological analysis. The writer also highlighted the idea that some of the units used by Efendiyev can be exemplified as epic onomastics.

Şekip Aktay (*Rus Halk Masallarındaki Antroponimler Uzerine Bir İnceleme*) analyzed anthroponyms used in Russian folk tales on the assumption that anthroponyms may provide valuable information about the community’s history, as

well as its political and economic structure. Moreover, he also stressed that names may enclose significant evidence about a community's worldview, its structure, culture, language and traditions. According to the writer, this might be the reason why personal names have been focused on by researchers who study sociolinguistics, linguistics, history of culture and ethnography. It was also mentioned that names are not limited with their vocabulary meanings only since they contain connotative meanings charged by the community culture. This implies that names contain religious, national, historical or mythological connotations. The author expressed his belief that from this perspective, names used in folk tales provide valuable data to analyze social acceptance and social values of the community.

Dealing with the origins of folktales, Aktay mentioned three of them: home culture, myths and imagination of the storyteller. It was also mentioned that heroes described in tales could be both imaginative and real people representing a specific stratum.

According to the author, anthroponyms in folk tales contain various semantic details and connotations and, therefore, they turn into linguistic elements with their semantic and symbolic meanings. In this framework Aktay analyzed both real names and imaginative names used in Russian folktales. It was mentioned that creators of folk tales use real names to express their attitudes towards the heroes. The writer supported this view by providing examples from Russian folk tales like *İvan-İvanuşka, Alyona-Alyonuşka* and *Sneg-Snegurocka.* According to the writer, the aim here is to cause to invoke the audience's sympathy to such heroes by using diminutive elements. Names given to heroes often collocate with words that express appearance (*Vasilisa Prekrasnaya – Guzel kız Vasilisa*), social staus (*Koylo cocuğu İvan, Kral İvan, Coban Frol*) or personality traits. As the world of tales contain fantastic and mystic elements, one can witness such imaginative names as Tekgoz (*Odnoglazka*), Olumsuz Koşcey (Koşcey Bessmertnıy), Morozko who possess extraordinary traits which differ them from real people.

Aktay reiterated the fact that *Ivan* is the most frequently used name in Slavic tales. In fact, this name can stand alone (*İvan, İvaşka)* and in various combinations

(*İvan Bıkovic, İvan Vodovic, İvan Kobıl'nikov, İvan Dorogokuplennıy, etc.*). The name *Ivan* is perceived in the West as a cliché to symbolize original Russian national features.

The writer spoke of the existence of such dominant names as *İvan-Durak* and *Car İvan* in Russian tales. Such names define the readers' attitudes to them. They also indicate whether they will win or lose, what kind of events they may face or, what their appearance or social staus is (*Vdovıy Sın, van Devkin Sın*, *İvan Kuharkin Sın*, *İvan Krest'yanskiy Sın*, *İvan Soldatskiy Sın*, *İvan Neschastnıy.)*. *Ivan* is trustworthy and usually wins his battles and therefore local people sympathize with him and award him.

Aktay analyzed the name *İvan- Durak* with emphasis on the etymology of the adjective *Durak* and the reasons for giving such a name with this adjective (*aptal*). The writer also focused on the characteristic features substantiated with this hero. The researcher also focused on the etymology of such images as *Baba- Yaga, Ejderha Gorınic, Olumsuz Koşcey, Leşiy*, *Vodyanoy* ve *Domovoy* and analyzed the features that are substantiated with them.

Aktay concluded that getting insights into the process of cultural development that personal names are used in helps understand the role of personal names in expressing the community's worldview. The writer spoke of the need to classify anthroponyms from semantic and structural perspectives and to reveal cultural, sociolinguistic, etymological and historical meanings.

2.2.2 Various approaches to dialect analysis

Ali Akar (*Ağız Araştırmalarında Artzamanlılık Sorunu*) first focused on two dialect compilations of Turkish carried out by native researchers stressing the fact that transcribed texts in one of them were not subjected to grammatical analysis. The researcher dealt with certain methodological problems both in formation and developmental processes of these structures. According to the researcher, the main problem was that Synchronic rather than Diachronic Grammar Method was used in the linguistic analysis of the developmental processes. It could be seen that such widespread alterations in the dialects as /g-/ < /k-/: *gapı* "kapı", /d-/ < /t-/: *dutmak*

"tutmak", */-ḫ, -ḫ-/ < /-ḳ, -ḳ-/*: *yoḫ* "yoh" or /o/ ve /u/: *böyümek* "büyümek" were compared with the standard variety of the Turkish language (*Türkiye Türkcesi*). But, there are different structures in dialects as well as common features with the written language. The researcher underscored the idea that the formation and development of these structures needed to be dealt with apart from written language. Moreover, dialect features needed to be considered vis-a-vis specific historical period.

Şınar Auelbekova *("Cul" Kökünden Türeyen Kelimelerin Kazakca ve Turkiye Turkcesi Ağızlarında Kullanımı ve Etimolojik Ma'na Kapsamı*) analyzed the etymology of the verb "*culga*" both in contemporary Kazakh and other dialects. The paper also focused on the use of the word "cul" and its derivatives and the main emphasis is placed on "kebenek" which can be the Turkish equivalent of the Kazakh "*culga*". The analysis was carried out from semantic and stylistic perspectives. The words that derive from Turkish "cul" which are used in the form of "*şulğau*", "*şulıq*"," "şulqa" in Kazakh were subjected to etymological meaning analysis. It was revealed that the word "kebenek" is used similarly in various Turkic varieties. Thus, an attempt was made to prove that the word "cul" in Turkish is the equivalent of the Kazakh "kebenek". It was also indicated that the word "kebenek" which means epidemic animal disease in Kazakh and Kirgiz can be taken as homonymous to the word "kebenek", the counterpart of the word "Cul". Emphasis was placed on stylistic analysis of proverbs that contain "Cul" and "kebenek".

Historical sources were used to get the data containing "cul" in various forms and the data were subjected to etymological analysis. The writer expressed her belief in the importance of such analysis since etymological analysis, according to the writer, may reveal fraternal ties and cultural closeness of Turkic nations.

Auelbekova spoke of the existence of words in Kazakh whose meanings are unknown to people. She expressed her concern that such words are likely to be lost in Kazakh. To support her view, she focused on the verb *"şulga"* which can be encountered in written samples from the past in Kazakh. In fact, the verb "*şulga"* is not used in contemporary Kazakh. The writer scanned a huge bulk of literary samples that have been created throughout Kazakh history. Emphasis was placed on the meanings expressed by this word in various forms. Having traced the origins and developments

of the word from structural and semantic perspectives, the writer concludes that Turkish "cul" and Kazakh "*culga*" are of the same origin. She also focused on the word and its derivatives in such varieties as Resey and Saratov dialects of Kazakh, Kirgiz, Tatar, Nogay, Tuva, Hakas, Yakut, Turkish, Azerbaijani varieties. An attempt was also made to analyze the semantic concept of the word "cul" used in proverbs in Turkish and its equivalents in Kazakh.

Auelbekova stated that although the word "cul" is not used in contemporary Kazakh, its derivatives as "cul+ğau" and "cul+qa" are still in use maintaining its historical origin and semantic meaning. She concluded that she intended to analyze such common words as "cul" and "culga" by tracking the historical process that they went through in terms of their structure and meaning. The writer also spoke about the need to analyze diachronically the mutual influences of various Turkic languages diachronically and the need to depict the words coming from the same origin and to compare them.

Huseyin Yıldız (*Türkiye Türkcesi Ağızlarında Alıntı Kelimelerde Gorulen /j/ Sesinin Kökeni ve Fonolojisi*) started his presentation with a generally accepted view that one of the criteria to define whether a word is borrowed or it is an original Turkish word is to look at the absence of the phone /j/. This implies that native Turkish words cannot contain /j/. As the writer mention, *Türkce* in the paper refers to Turkish spoken in Turkey. It was also mentioned that /j/ is frequently used in contemporary Turkic languages as Kazakh, Karakalpak and Tuva. Yıldız also underscored the fact that *Türkiye Türkcesi Ağızları Sozluğu* (*The Dictionary of Turkish Dialects of Turkey*) contains 226 words with this sound whereas the number of words with /j/ in *Sesli Türkce Sözlük* (*Turkish Voice Dictionary*) is 900.

Although Yıldız divided words with /j/ into two categories (Loan words and Native Turkish words), emphasis was placed on the loan words only. Then, the author dealt with the distribution of words with /j/ in accordance with donor languages: Arabic, Persian, French, Russian, Caucasian languages (Karachay < Circassian), Armenian, English and Italian. It was found out that among these seven languages, the number of words with /j/ is high in such languages as *Arabic, Persian and French.* According to Yıldız**,** this can be observed in two ways: preservation of the sound */j/* →

[j] and transformation of words with various sounds to words with */j/* in Turkish dialects*:* → ? > j. The author focused on each of these ways and provided specific examples.

It was reported that among the words in which /j/ preserved its properties the number of original French words significantly prevail (out of 47 cases 43 are originally French) while 2 of them originated from English, 1 from Italian and 1 from Persian. Concerning their positions, it was observed that /j/ can be seen in four positions in French while in Italian it is in the initial position only. It was also found out that /j/ is in the final position in the word that originated from Persian. Generally speaking, the number of words where /j/ preserved its properties prevailed in words that were borrowed from European languages.

The writer also mentioned the fact that the sound /j/ has undergone alterations with such sounds as /c/, /ş/ and /z/ in two perspectives. It was reported that *j ~ c* alterations in the direction of *j < c* was observed only in the words of Arabic and Persian origin while *c < j* direction prevailed in words of French and English origin. There was only one case with a word from the Persian language.

It was also reported that *j ~ c* alterations in the direction of *j < c* were observed in words that originated from Russian, whereas *c < j* direction was more frequent and these words, as a rule, were of French origin. The author mentioned that *j ~ ş* alterations in the direction of *j < ş* were observed in three positions excluding in initial positions while *ş < j* direction was observed in final positions in words that originated from French. Finally, *j ~ z* alterations in the direction of *j < z* were observed in the words of Arabic origin in the final position and the alteration was observed in initial position in a Persian originated word. *z < j* direction alteration was observed in one Persian origin word in final position. Yıldız reported that on cases when /c/, /g/, /s/, /ş/ and /z/ in donor languages can transform into /j/ in Turkish dialects in Turkey. It was found out that with one exception (Russian) all these words come from Arabic.

Gonca Demir (*Türk Halk Muziği Fonetik Notasyon Sistemi/THMFNS Jeolekt-Muzikolekt Ozellikleri: Urfa Yöresi Orneklemi*) focused on the geolectic and musicolectic properties of the folk music specific to the Urfa area of Turkey by using Turkish Folk Music Phonetic Notation System (THMFNS). The aim here seems to

initiate a parallel application to the international linguistic/musicological application foundations and to develop it in configuration with phonetics, morphology and lexicon in the axis of traditional/international attachments based on Standard Turkish and International Phonetic Alphabet sounds.

2.2.3 Dialect analysis from various perspectives

Sonnur Aktay (*Toplum–Kültür–Dil Etkileşimlerinin Yer Adlarına Yansıması*) viewed place names as objects to reflect a nation's world view, its traditions, culture, and the link between its past and present. According to the author, as place names contain too much covered or uncovered information, they are focused on by researchers of various fields like geography, sociology, history, anthropology, archeology, linguistics, etc. According to the writer, place names are seen as one of the main sources in the study of language history, dialectology and etymology since place names of a certain region are directly associated with the people of that region, their history and language variety. Therefore, when analyzing place names, it is important that the researchers refer to geography, history, sociology, biology, archeology and anthropology.

Stressing the importance of the etymological analysis, the author expressed her belief that systematic and multilateral approach used in etymological studies enables the researcher to reveal the place names' origin, history, real meaning, contact language culture since place names contain features that come various communities, cultures and languages. To support this view, the author focused on place names of Turkic origin in various parts of Russia (*"Ulan-şibir", "Karaşibir", Астрахан, Сарамов, etc.).*

The writer also touched upon the importance of history, language, politics and national consciousness. By providing specific examples, the author stressed the importance of taking such elements as the previous name's connotation, its history, etymology, social, political, economic and cultural developments when naming or renaming a street or a square into account. It was also mentioned that in order to form national consciousness and impose the intended ideology, authorities permanently change names. The writer provided examples of place names which reflect various

important events in Turkey like *Kurtuluş Caddesi, Demokrasi Caddesi, Cumhuriyet Meydanı, İstiklal Mahallesi, Hürriyet Mahallesi, etc.*

The main emphasis was placed on Russian place names vis-à-vis socio-political changes in Russia and in various post-Soviet republics. Comparing the reasons of change in Tsarist, Soviet and post-Soviet periods, the author stressed that street names given during the Tsarist period underwent changes during the Soviet period while not all the names given the Soviet period were changed. Moreover, street names of the Tsarist period did not have any ideological connotation, but rather they had referential meaning (they were used to provide information). However, Soviet period names were given by ideological considerations. What concerns post-Soviet period names, national or nationalistic considerations prevail.

It was also mentioned that focusing on old maps in the analysis of place names widens the scope of such studies. The author expressed her belief that in this way one may get better insights into historical events, inhabitants and their cultural values since place names contain traces of historical events, culture, contact languages and the people's way of life.

Serdar Bulut (*Türkiye Türkcesi Ağızları İle Kıbrıs Türk Ağızlarında Ortak Kullanılan Kalıp Sözlerden Hayır-Dualar Ve Beddualar*) focused on such formulaic expressions as benedictions and maledictions from the comparative perspective mentioning that, although mainland Turkish and Cypriot Turkish share many formulaic expressions, there are also expressions that are specific for Cypriot Turkish. Therefore, the author aimed to classify formulaic expressions in accordance with the field of use both in mainland Turkish and Cypriot Turkish. While doing this, emphasis was placed on the possible differences in terms of phonology and form. Rather than carrying out a fieldwork, the author seems to prefer using printed materials on the subject. The author considered formulaic expressions or cliché words to reflect the material-spiritual cultural values and beliefs of the society. They are usually charged with original, impressive and emotional connotations. Moreover, they contain high affective content; they encompass the user's attitudes which may include joy, gratitude, sorrow, anger and hatred. Patterns of benediction, curses, greetings, oaths and others are the ones encountered in everyday life. In fact, all regional dialects of Turkish have their

own specific 'mold' patterns. These 'mold' words are used in many different ways in the dialects of Turkey Turkish and Turkish Cypriot.

Bulut focused on the topic from continuity and discreteness perspectives without using these terms. In other words, the author spoke of both similarities and differences in using benedictions and maledictions. It was reported on the fact that there are many similar words which carry the same meaning in both dialects of Turkey Turkish and Turkish Cypriot. It was also found out that patterns or words which are used in Cypriot Turkish are non-existent other dialects. Likewise, the words witnessed in some 'mold' expressions in mainland Turkish dialects were not observed in Cypriot Turkish dialects. The presenter contemplated on benedictions and maledictions which are commonly used in *Türkiye Türkcesi* and Cypriot Turkish in accordance with the field that they are used: *marriage, birth, death, health, taking food,* or wishing someone's death, bad luck, misfortune, starvation, etc. While doing this, emphasis was placed on sound and shape differences.

It was found out that although *Turkiye Turkcesi* and Cypriot Turkish dialect have many common expressions, there are also expressions used in Cypriot Turkish only. This view was supported by specific examples. Concerning the linguistic features of formulaic expressions, the writer mentioned the use of imperatives, particles as *"-asıca, -esice*", the word "*İnşallah*" and some phonological elements. It was also mentioned that emphasis was placed on common features.

Fatih Kurtulmuş (*Turkiye Turkcesi Ağızları Uzerine Yapılmış Calışmalarda Unluler ile İlgili Problemler*) started his presentation with an insight into the history of research on Turkish dialects which date back to 1867 and divided it into two periods: studies carried out by non-Turkish researchers (1860s and 1940s) and the ones carried out by Turkish researchers (from 1940s till now). Then, the author dealt with two compilations carried out by Turkish Language Society before 1960. However, the main emphasis was placed on the problems that could be seen in the studies concerning vowels.

According to Kurtulmuş, one of the challenging issues that researchers have long been engaged in is letter-word discrepancy in writing systems. In this regard, the researcher touched upon two perspectives. 1) Indicating the same sound with two

letters and symbols. 2) Using the same letter and symbol to indicate various sounds. But this view has not been accepted unanimously by researchers. It is believed that stereotyping and standardization can be the best approach to take.

The writer explained what is meant by stereotyping (converting texts with various writing systems to a system accepted by everyone) and standardization (to present the stereotyped writing system to researchers by using a writing system like UNICODE). Kurtulmuş expressed his belief that the problem with research on dialects can be solved if standardization like Unicode had been carried out.

Kurtulmuş also dealt with the writing type, characters and fonts to be used. He explained how stereotyping and standardization can be achieved. It was also mentioned that writing types like Gentium and Oktay New Transcription used in studies by Turkologists have deficiencies. The presenter suggested that a different keyboard application "UKELELE" be used to deal with this deficiency. The presenter also touched upon the main features and advantages of the proposed keyboard system.

According to the author, in order to fully realize phonetic properties of Turkish dialects in Turkey, there is a need to solve the problems concerning the definition of vowels and their environment. The author expressed his belief that in parallel to the technological developments nowadays, it would be possible to transfer dialect samples to computer programs and work out data processing programs which may help to follow the role of the environment in allophone alterations.

The author expressed his belief that technology use in research studies is not at the required level. Although the need for phonetic laboratories has long been touched upon, laboratories are not used.

Kurtulmuş also spoke about the need to follow the recent developments taking place in computational sciences, especially in the field of "Natural Language Processing" in dialect studies. He expressed his belief that some of the problems mentioned above could be solved through integrating technology with research on dialects.

Sezer Ozyaşamış Şakar (*Türkiye Türkcesi Ağızlarında Ötümsüzleşme*) touched upon the phenomenon of devoicing and focused on both native Turkish and loan

words in Turkish dialects in Turkey stressing that, unlike standard Turkish words, devoicing can be witnessed in Turkish dialects in initial, middle and final positions. The author also touched upon the reasons for devoicing. Dealing with the scope, Şakar mentioned that the geography of the dialects is not limited within the national borders of Turkey. In this framework, it would be impossible to analyze all Turkish dialects in terms of devoicing. Therefore, the aim in the paper was to scrutinize various materials including books, PhD dissertations, MA theses and articles on dialect analysis to describe words in which the phenomenon of takes place, to see whether these dialects possess their own discrete properties, to define the influence of historical and ethnic elements in the phenomenon of devoicing in Turkish dialects.

Şakar analyzed the phenomenon of devoicing in various dialects by providing specific examples. Among the alterations mentioned in the paper are: *b > p, d > t, g > k, ġ > ḳ, ğ > ḫ, j > ş, v > f, z > s.* It was mentioned that the phenomenon of devoicing occurs in in Turkish dialects in Turkey and such alterations cover the following phonemes: b > p, d > t, g > k, ġ > ḳ, ğ > ḫ, j > ş, v > f, z > s. The author also mentioned that the frequency of such alterations is not the same in all dialects.

According to the presenter, there can be several sources of such alterations. It was reported that certain words preserve the original old Turkish forms while in certain native Turkish or in loan words this happens under the influence of the voiceless sounds. This may also happen due to the historical fact that certain non-Oghuz groups settled in the territory which covers North-eastern dialects. Among such reasons the writer also mentions the role of idiolects.

Seval Dirik (*Çanakkale Ağızlarında Bir Kiplik Şekli: "-mAk vā"*) expressed her belief that compared with other dialects, Western Group dialects had not been analyzed intensively and therefore it would not be quite unexpected if one encountered linguistic elements different from other dialects. In this vein, Dirik contended with a pattern that had never been touched upon from the perspective. Thus, the presenter first provided information about modality in Turkish mentioning that in standard Turkish, when the speaker wants to express he wishes something to happen, he uses pattern like "*Şimdi Ankara'ya gitmek var*" where the word "*var*", which means "*possession*", grammaticalizes and expresses the action that the speaker wants to

happen. In such cases the intended subject is the first person singular. However, as was reported, the same sentence with quite a different semantic meaning can be seen in Sazlı Koyu dialect in Canakkale area: *Şimdi Ankara'ya gitmek vā* which implies that *he must be on his way to Ankara now*. According to the reporter, unlike the properties of the standard language, the sentence in Sazlı Koy dialect expresses speculation, deduction and probability. The subject of the sentence refers not to the first person singular, but to the third person singular. In other words, it implies the speaker's perception of the action carried out by the third person.

Thus, Dirik focused on functions and meanings of "*-mAk vā*" encountered in Sazlı Koyu dialect in Canakkale area. She touched upon the functions and semantic properties of the pattern with "*-mAk vā*" meticulously and indicated that it should be taken as epistemic modality expressing speculation, deduction and probability. In other words, it can be thought that, in this respect, the semantic and stylistic features of the pattern that emerge as a result the grammaticalization of the word "*var*" presents a different view to the concept of modality. The paper also touched upon the morphological properties of the pattern: "*-mAk vā*". It was also noted that the pattern is not used in interrogative sentences.

2.3 Conclusion

As could be seen from the synopses of the papers presented to the conference, researchers addressed a wide range of issues concerning dialectology and geolinguistics from various perspectives. Both the depth of analysis and innovative approach to the field enquiry contributed to the success of the congress. Originality of many papers and distinctiveness of the findings can be mentioned among the main assets. Enthusiasm and high professionality of the organizers, uniqueness of the venue and high interest of the audience were also among the factors that made the 8th Congress of the International Society for Dialectology and Geolinguistics memorable.

Dialectologia 17 *(2016), 227-238.*
ISSN: 2013-2247

FIRST DIALECTOLOGISTS

FRAN RAMOVŠ

(1890-1952)

"The linguistic atlas is an indispensable means for the study of dialect and language. […] By this we do not mean a collection of colourful maps displaying the range of this or that dialect; it is not a linguistic atlas but a dialectological map, an approximate projection of a certain number of linguistic phenomena, the range of which almost never completely corresponds to the range projected in this way. [...] A linguistic atlas, however, contains a larger number of maps; each map shows/ demonstrates only one phenomenon, one word, its realization in speech

(almost) simultaneously across the whole Slovene territory; the atlas is thus a sort of dictionary, which is not alphabetized but geographically arranged. [...]

A collector should always be, i.e. in all places one and the same person; this person, who perceives with the same ear the Slovenian sounds in all the different places, will use the same transcription technique also in jotting them down. This person should be well educated in phonetic transcription and should possess a well-trained ear; it is important that such person distinguish qualitative, even ever so tiny differences, in order to hear the intonation of unstressed syllables, and, naturally, also the accentual nature of stressed syllables. Even if I owe it to coincidence, I am fortunate to have such a person" (Ramovš 1934: 1-4).

1. Biography

1890 – born in Ljubljana on 14 September.

1910-1940 – studies linguistics first in Vienna, where his teachers were Mayer-Lübke and Kretschmer, then in Graz, where his head teacher was R. Meringer.

1911 – becomes substitute assistant to Prof. Meringer and starts writing his doctoral thesis.

1912 – finishes his doctoral thesis on the development of Proto-Slavic semivowels in Slovene. In the same year he starts preparing his habilitation thesis on modern consonant reduction in Slovene.

1914 – hands in his doctoral thesis and is promoted to Doctor of Philosophy in Graz. In the same year he also completes his habilitation thesis entitled *Slovenische Studien*, and also publishes his first work on Slovene dialects entitled *Zur slovenischen Dialektforschung*.

1914-1936 – publishes works on Slovene dialects.

1915 – mobilized into the army and sent to the Isonzo front. Two years later he is released from active army duty due to impairments.

1918 – obtains a habilitation at the University of Graz; following the disintegration of the Austro-Hungarian Monarchy he returns to Ljubljana, where he is elected secretary of the university commission at the University of Ljubljana.

1919 – following the founding of the University of Ljubljana he is one of the first to become Professor of Indo-European and Slavic language studies.

1921 – becomes secretary of the Scientific Society for Humanities, and five years later the dean of the Faculty of Arts at the University of Ljubljana.

1934 – becomes rector of the University of Ljubljana; he proposes his plan for the founding of the Slovene Academy of Sciences and Arts and prepares a draft for the Slovene Linguistic Atlas.

1938 – as the Slovene Academy of Sciences and Arts (SAZU) is founded he becomes its full-time member.

1945-1950 – is the main secretary of SAZU, and its president from 1950 to the time of his death.

1945-1952 – leads the Institute for the Slovenian language at SAZU and prepares a long-term work plan which the Institute, named after him since 1986, still follows in most points even today.

1952 – dies in Ljubljana.

2. Ramovš's main works

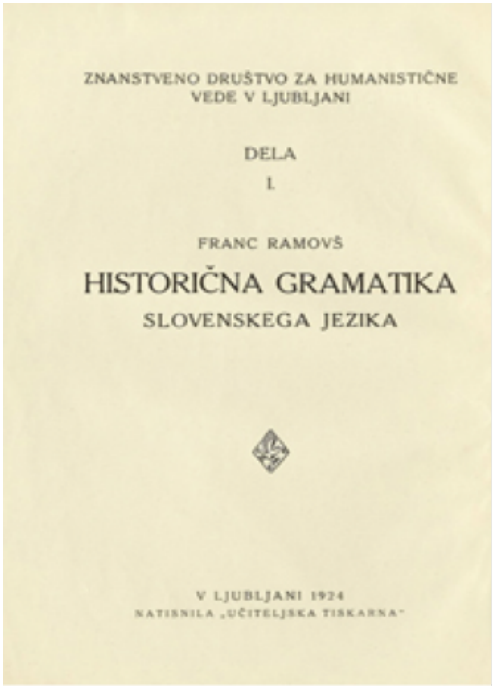
1914 – *Zur slovenischen Dialektforschung*, AslPh,[1] pp. 329-337. This is Ramovš's first paper on Slovene dialects which focuses on the development of individual linguistic phenomena observed both in the dialects and in the standard variety.

1918, 1920 – *Slovenische Studien*, AslPh 37, pp. 123-174; pp. 289-330. Ramovš's habilitation thesis dealing with the modern vocalic reduction in the Slovene language, printed in two volumes. The materials for his research on modern vocalic reduction are

[1] AslPh – *Archiv für slawische Philologie*.

his own excerpts from Slovene manuscripts and from the majority of Slovene books dating from the 16th century onwards, and materials from Slovene dialects.

1924 – *Historična gramatika II: Konzonantizem (Historical Grammar II: Consonantism)*. Ljubljana, 335 pages. This was Ramovš's first book from the scheduled series Historical Grammar of the Slovene Language. Even today it is a reliable and abundant source for all phenomena concerning Slovene consonantism and its thousand-year historical development as we know it from Slovene written and printed sources and dialects.


ZNANSTVENO DRUŠTVO ZA HUMANISTIČNE VEDE V LJUBLJANI

DELA

I.

FRANC RAMOVŠ

HISTORIČNA GRAMATIKA

SLOVENSKEGA JEZIKA

V LJUBLJANI 1924

NATISNILA „UČITELJSKA TISKARNA"

Historična gramatika slovenskega jezika II – Konzonantizem (Historical Grammar of the Slovene language II: Consonantism)

http://www.fran.si/slovnice-in-pravopisi/38/1924-1952-ramovs

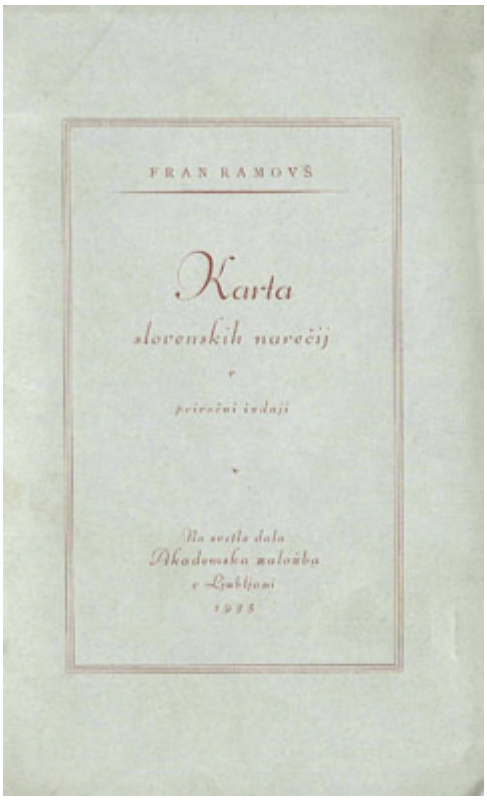
1928 – *Die Charakteristik des slovenischen Dialektes in Resia*, ČJKZ 7,[2] pp. 107-121. Ramovš's first study which offers a reliable insight into the development of a Slovene dialect in its entirety.

[2] ČJKZ – *Časopis za slovenski jezik, književnost in zgodovino*.

Dialectologia 17 *(2016), 227-238.*
ISSN: 2013-2247

1929 – *Slovenački jezik (Slovene Language)*, Narodna enciklopedija srpsko-hrvatsko-slovenačka, volume 4, pp. 192-208. In this paper Ramovš for the first time gives a more in-depth account of the entire Slovene language dialectal area.

1931 – *Dialektološka karta slovenskega jezika (Dialectological Map of the Slovenian Language).* Ljubljana, 72 pages + map. In this book Ramovš publishes the first scientifically based map of the Slovene dialects, divided into 7 dialect groups with more than 40 dialects. The map has to date required modifications only in places for which the author lacked research material.



Dialectological Map of the Slovene Language
http://www.narecna-bera.si/koroska-narecja/slovenska-narecja

1933 – *Kratka karakteristika slovenskega narečja na Dolenjskem (A Short Survey of the Slovenian Dialect in Lower Carniola)*, Sbornik Miletič, pp. 153-163. This paper reflects the final development of his method on dealing with dialects: the description of the borders of the discussed dialect is followed by the general acoustic impression and the development of reflexes of individual Proto-Slavic vowels. There is also an

account of the influence of consonants on individual vowels (positional development). He continues with consonants and morphology, and concludes with the specifics of the dialect or speech in question.

1935 – *Historična gramatika VII: Dialekti (Historical Grammar VII: Dialects)*. Ljubljana, 204 pages + map. Ramovš's second book from the series Historical grammar of the Slovene language which presents the dialectal diversity of the Slovene language and the descriptions of individual dialects is even today the fundamental synthetic work on Slovene dialectology. In it Ramovš makes a synthesis of the entire contemporary knowledge of the Slovene dialects, clearly and conclusively divided into 7 dialect groups with hardly any need for corrections.

1936 – *Kratka zgodovina slovenskega jezika I (A Short History of the Slovenian Language I)*. Ljubljana, 246 pages. In this book Ramovš presents the development of all Slovene vowels originating from Proto-Slavic ones in long and short syllables. It replaced the third book which was scheduled for the series Historical grammar of the Slovene language entitled *Vocalism*. He never wrote the latter due to extensive teaching and scientific-organizational obligations and to poor health. According to experts this is his best book in which he makes an overall account of complex Slovene vocalism.

1937 – Together with M. Kos he prepares a new edition of *Brižinski spomeniki (Freising Manuscripts*), the oldest known and preserved records of the Slovene language and the oldest text in Latin script in any Slavic language. In the phonetic transcript of this text Ramovš defined the entire phonetic image of the Freising Manuscript, which has to date required merely minor adjustments. With the linguistic analysis of this text he proved that it was undoubtedly written in the Slovene language and that the transition between the Proto-Slavic dialect into the Slovene language was complete as early as in the mid-8th century.

1950 – *Relativna kronologija slovenskih akcentskih pojavov (Relative Chronology of Slovene Accentual Phenomena)*, Slavistična revija III, pp. 16-23. Ramovš did research also in the Slovene development of stress and pitch accents; he defined and explained accent shifts in Slovene, their order and geographic boundaries.

1952 – *Morfologija slovenskega jezika (Morphology of the Slovene Language)*. Ljubljana, 167 pages. Due to poor health Ramovš was unable to prepare this book on the history of Slovene morphology himself. His students had collected notes from his lectures before his death. Ramovš, his health severely deteriorating, approved its being published.

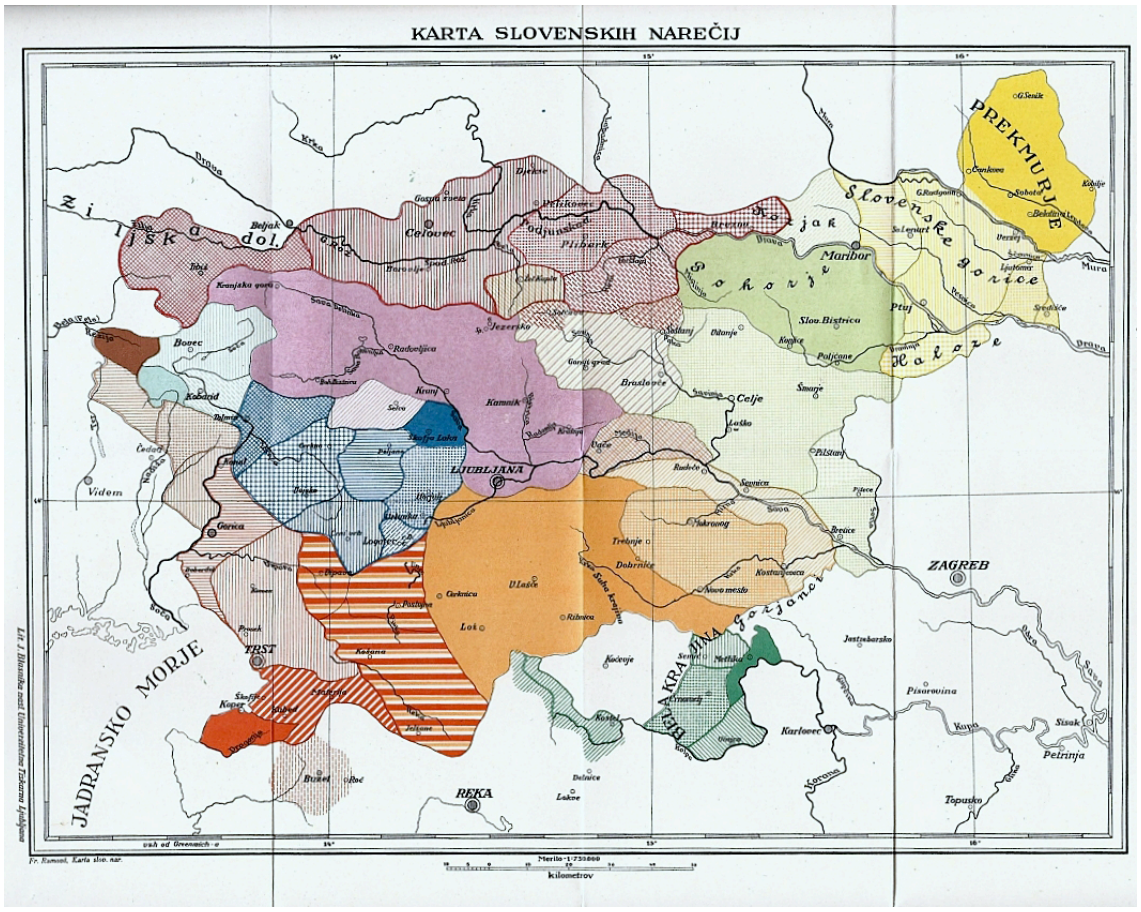
3. Ramovš's work

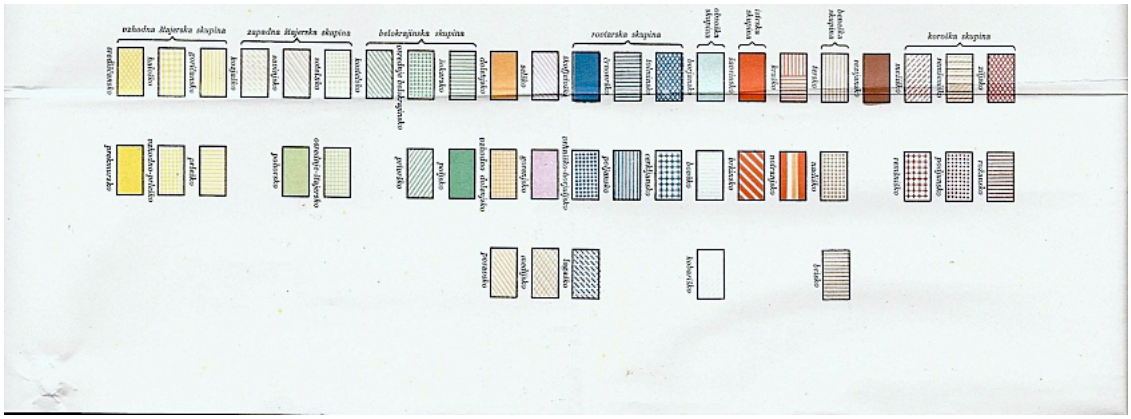
Fran Ramovš is considered one of the most significant Slovene linguistic scientists. He began his scientific career when larger Slavic nations had already started working on major synthetic works, while in Slovenia the interest in language seemed to fade with the oncoming First World War. With the exception of some random dialectological seminar papers and dissertations from the Universities of Graz and Vienna, nothing new was written about the Slovene language. As Ramovš began his working career he found himself in a vast untended field which still lacked even the majority of preparation studies. He had emerged from the school of Neogrammarians firmly anchored in the framework of Indo-European comparative grammar, whose views and methods would soon not be of any use to him. He sought new paths by familiarizing with Schuchard, one of the first linguists to determine the principles of sound substitutions, and with the linguistic views of Meillet, Bartoli, de Saussure, Sievers, Jespersen, etc. He applied all the principles of modern linguistics in his study for a detailed depiction of the Slovene language, and in doing this he also invented the entire necessary Slovene linguistic terminology. Already as a student he systematically and methodically studied all the fundamental Slovene texts up to the mid-19th century and at the same time excerpted them for all the chapters of the historical grammar of the Slovene language, especially for phonology and morphology. In addition, he was also profoundly involved in the study of live Slovene speech, and in the course of doing fieldwork (he roamed a significant part of the Slovene language territory) or with the help of interviewees he familiarized with almost all the Slovene dialects. These had up

to then been fairly unknown, and the that time classifications did not provide a realistic picture of the Slovene dialectological complexity and interrelatedness.

Soon after writing his first disquisitions, Fran Ramovš proves himself to be an autonomous connoisseur of linguistic phenomena in the Slovene literary language and dialects, and it does not come as a surprise that already in the introduction to his habilitation thesis entitled *Slovenische Studien*, where based on materials from older Slovene literature and dialects he described the newest general Slovene linguistic phenomenon, namely the modern vocalic reduction, he announced the emergence of his Slovene historical grammar. In his book *Konzonantizem (Consonantism)* (1924), the first book in this series to be prepared for publishing, Ramovš solved almost every single problem relating to Slovene consonantism. Even when it comes to the current situation in Slovene dialectology and historical grammar, this book describes, explains and temporally and spatially determines almost every phenomenon concerning the Slovene consonantal system. The relatively conservative nature of Slovene consonantism did not cause any particular difficulties for Ramovš, which is why he devoted much more effort and time to the complex Slovene vowel system (vocalism). The deeper he reached into the studies of Slovene dialects, the more questions emerged, since he was not satisfied merely with establishing the synchronic state of the Slovene dialects. He wanted more, namely he wanted to connect the dialectological material with its historical development. First he completed his book *Dialektološka karta slovenskega jezika (Dialectological Map of the Slovene Language)* (1931). Based on several external and internal linguistic factors this book represents the first scientifically supported classification of the Slovene dialects with its seven dialect groups. Since it emerged it has needed no significant corrections.

Dialectologia 17 *(2016), 227-238.*
ISSN: 2013-2247





Four years later Ramovš presented the final definition of his dialectological views in the book *Historična gramatika VII: Dialekti (Historical Grammar VII: Dialects)* (1935). In *Dialekti (Dialects)*, which Logar (1991: 15) labelled the golden book of Slovene studies, Ramovš presents the description of all the dialects and provides their classification. In this classification he accounts for over a thousand linguistic data alongside that regarding population, geography, colonization, society, state

administration and church administration. Due to the latter factors, the Proto-Slovenian (or Alpine Slavic) language, which had been until the 12th century relatively unified, developed into today's extremely variegated dialectological mosaic. Ramovš divided the Slovene dialects into 7 dialect groups with more than 40 dialects and speeches. Since this division, only few corrections have been made and supplements added on the basis of new findings (e.g. those regarding dialect boundaries). In *Kratka zgodovina slovenskega jezika (Short History of the Slovene Language)* (1936), which was published soon after *Dialects* and by Logar (1991: 16) described as a true masterpiece in Ramovš's scientific opus, Ramovš concisely and systematically describes the autonomous development of the Slovene language from its earliest times till today. Logar notes (ibid) that by accounting for substitutional, phonetic, phonological, psychological, historical, geographical, sociological and other factors Ramovš was able to successfully connect dialectological reflexes with the historical development of sounds. He developed a genealogy of reflexes in all long vowels in the Slovene dialects which clearly shows their development from the Proto-Slavic language onwards. He also illustrated it with sketches of the Slovene linguistic territory, in which individual reflexes and their geographical distribution are demonstrated for each vowel separately. Ramovš was aware of the fact that this book does not solve all the problems of Slovene vocalism, since the latter is extremely complicated due to the close association between accent and vowel quality and due to the large number of dialects. Just how complicated Slovene vocalism is can be seen in his studies *Relativna kronologija slovenskih akcentskih pojavov* (*Relative Chronology of Slovene Accentual Phenomena* (SR III, 1950) and *Osnovna črta v oblikovanju slovenskega vokalizma* (*A Survey on the Formation of Slovene Vocalism*) (SR IV, 1951), in which he clearly fixes the dependency principles of vowel quality on vowel quantity, this being one of the fundamental principles in the development of Slovene vocalism. In the course of developing the dialectological map, Fran Ramovš also came up with the idea of a Slovene linguistic atlas (SLA), and through meticulous study of dialects saw to its development. This fundamental work of modern Slovene dialectology and geolinguistics, for which he had prepared a questionnaire with a designated network of 312 places – net points, was designed in 1934. The preparations for the SLA at the

Institute for Slovene Language at the Slovene Academy of Sciences and Arts (SAZU) began after the Second World War, and the first volume of the SLA entitled *Človek (Man)* with questions and sub-questions relating to *body, diseases, family,* came out in Slovenia as late as in 2011.

Ramovš did not particularly fancy the standard variety of Slovene, however, he did tackle this issue for the first time before the war and together with A. Breznik published a Slovene normative guide (1935). The second time was after the war when the Institute for Slovene Language with the SAZU took over the task to publish a new Slovene normative guide. Despite his illness he supervised the making of the guide to its end in 1950. In 1942 he presented the draft for an etymological dictionary of the Slovene language, and in 1946 also the layout for the collection of materials for the extended standard Slovene dictionary. The latter was the basis for the later published *Slovar slovenskega knjižnega jezika I–V* (*Standard Slovene Dictionary I-V*) (1970-91). He was the editor and co-editor of numerous scientific publications (e.g. ČJKZ, Južnoslovanski filolog, Razprave SAZU, LSAZU, Slavistična revija), while his scientific achievements won him the election of corresponding member of the School of Slavonic and East European Studies (1925), corresponding member of the Yugoslav Academy of Sciences and Arts in Zagreb (1926), of the Serbian Academy of Sciences and Arts in Belgrade (1929), the Slovenian Institute in Prague (1929), the Polish Academy of Science in Krakow (1935) and the Modern Linguistic Association in America (1948).

Ramovš dedicated his entire life to the Slovene language. He embraced his mother tongue to such an extent and depth that no later fundamental work in Slavic studies could avoid considering his findings. With in-depth studies with "innovative, convincing and fluent solutions based on masses of meticulously analysed and inventively connected documentary material" (Logar 1996: 78). Ramovš is deservedly appreciated as one of the most important Slovene linguists. Let me conclude with a thought by the most renowned Slovene etymologist Prof. Dr. F. Bezlaj who said, "For Slovenes, Ramovš is the beginning and end of everything we know about our language. There has been no one to date who would be so deeply devoted to the subject of

language development, save for Ramovš who even developed an entire synthesis for us literally out of nothing." (Bezlaj 1950: 225)

Mihaela KOLETNIK
Univerza v Mariboru
Filozofska fakulteta
Oddelek za slovanske jezike in književnosti, Koroška cesta 160
SI – 2000 Maribor
mihaela.koletnik@um.si

References

BENEDIK, Francka (2000) *Vodnik po zbirki narečnega gradiva za Slovenski lingvistični atlas (SLA)*, Ljubljana: Založba ZRC, ZRC SAZU.

BEZLAJ, France (1950) "Fran Ramovš. Pogled na njegovo delo", *Slavistična revija*, III/3-4, 225-236.

KOLARIČ, Rudolf (1960-1971) "Ramovš Fran", in A. Gspan (ed.), *Slovenski biografski leksikon*, Ljubljana: SAZU, 22-24.

LOGAR, Tine (1961) "Dr. Fran Ramovš in njegovo delo", *Jezik in slovstvo*, 7/3, 65-69.

LOGAR, Tine (1991) "Pomen Frana Ramovša za slovenistiko", in M. Orožen & I. Orel Pogačnik (eds.), *Sedemdeset let slovenske slovenistike*, Ljubljana: Zavod Republike Slovenije za šolstvo in šport, 13-17.

LOGAR, Tine (1996) "Ramovš, Fran", in A. Dermastia (ed.), *Enciklopedija Slovenije*, *10*, Ljubljana: Mladinska knjiga, 77-79.

RAMOVŠ, Fran (1934) "SLA – Slovenski lingvistični atlas" (typescript).

RAMOVŠ, Fran (1971) "*Zbrano delo. Prva knjiga*", T. Logar, J. Rigler (eds.), Ljubljana: SAZU.

RAMOVŠ, Fran (1997) "*Zbrano delo. Druga knjiga*", J. Toporišič (ed.), Ljubljana: SAZU.

TOPORIŠIČ, Jože (1987) *Portreti, razgledi, presoje*, Maribor: Založba Obzorja.

TOPORIŠIČ, Jože (1994) "Fran Ramovš kot narečjeslovec", *Slavistična revija*, 42/2-3, 159-170.