# THE DEVELOPMENT OF NEW CORPORA FOR UNDER-RESOURCED LANGUAGES USING DATA AVAILABLE FOR WELL-RESOURCED ONES

*Pavel Skrelin[1], Nina Volskaya[1], Karina Evgrafova[1], Riikka Ullakonoja[2]*

[1] Saint-Petersburg State University, Russia
[2] University of Jyväskylä, Finland

## ABSTRACT

In the paper we propose to exploit existing corpora of well-resourced languages as a basis for developing similar corpora of under-resourced ones. The construction of this type of corpora will allow finding common patterns of acoustic manifestation of similar functional states regardless of the language. The analysis of these corpora will also allow investigating universal and language-specific features reflected in speech. Two pilot experiments which may contribute to the proposed strategy are presented.

***Index Terms***— under-resourced languages, parallel speech corpora, acoustics, intonation

## 1. INTRODUCTION

Nowadays there are about 7000 human languages spoken in the world[1]. However, only about 20-30 of them have a significant digital presence. Such languages as English, Spanish, Mandarin, German, French and Japanese, for example, are characterized by information technology available, substantial presence on the Internet, existing software adapted for their use [6]. These and the next 270-280 most widely-spoken languages account for over 90% of the world's speakers; the other 10% or so of the population speaks one of 6700 minority languages. The problem of availability of linguistic recourses such as written or speech annotated corpora, ASR and TTS applications etc., is especially relevant for these languages also referred to as pi-languages or poorly-equipped languages. This concerns most or all the languages of the less developed countries of the world.

A substantial number of papers discuss strategies for improving strategies for under-resourced languages. [2], [6], [8], [9]. They often focus on collecting parallel and/or comparable corpora. The term parallel corpora refers to speech samples of parallel texts which are true and accurate translations or approximate translations with minor language-specific variations. Typically, these are translations of canonical texts, legal documents, fiction translations, software manuals (e.g. the Bible, the Universal Declaration of Human Rights etc.)[2]. By comparable corpora the samples of closely related texts reporting the same event or describing the same subject are meant. The parallel and comparable speech corpora are supposed to provide data on many languages with the same recording quality (sampling rate, microphone type, noise conditions), speaking styles (read, conversational), transcription and dictionary formats, and of the same domains. Today there seems to be a lack of big-size corpora of this type mostly due to fact that their construction involves a lot of funds. The existing ones provide us with mostly lexical, morphological etc., information.

We propose another category of parallel/comparable corpora which is multilingual parallel corpora of functional states (e.g. emotional states, fatigue, depression etc). They are supposed to be collected according to same scenario which triggers the verbal expression of a certain state.

The construction of this type of corpora will allow finding out common patterns of acoustic manifestation of similar functional states regardless of the language. The analysis of these corpora will also allow investigating universal and language-specific features relevant to different functional states and the ways they are reflected in speech.

Moreover, the main advantage of the approach is a possibility to exploit the existing corpora of well-resourced languages as a basis for developing similar corpora of the under-resourced ones.

For example, there does not exist a corpus of emotional children's speech for the automatic recognition of emotion related user states for Russian. However, there is one for German [1]. It provides acoustic data on emotion reflections in the speech of German children. Conducting the perception tests on the emotion recognition of this material by Russian listeners, we can find out similar and different ways of expressing emotions by speakers of both languages. Thus, a parallel corpus of emotion related states for Russian can be designed to include only the acoustic data which are specific to the language. As a result, the amount of effort and costs is minimized.

This approach provides also a good ground for testing our knowledge on emotion recognition patterns in the native speakers of various languages.

[1] http://www.ethnologue.com/

[2] http://rosettaproject.org/

The emotion-related features and stylistic characteristics of speech are defined by the intonation means. Thus, to collect parallel multilingual corpora it is necessary to find out universal and language-specific intonation patterns of expressing various functional states and pragmatic meanings.

In the paper two experiments which may contribute to the strategy described above are presented.

## 2. PERCEPTION OF EMOTIONS IN THE SPEECH OF GERMAN CHILDREN BY RUSSIAN SPEAKERS

The goal of this research was to study universal and language-specific means of expressing and perceiving emotions in the speech of Russian and German natives. The study was based on the speech material of Aibo Emotion Corpus that was kindly provided to us by our German colleagues [1]. The corpus included the speech of preschool children. Children's emotional expression is spontaneous as their behavior is practically not determined by social conventions. That is why the emotional children's speech is suitable for investigating direct correlation between acoustic characteristics of speech and emotional verbal reactions.

A pilot perceptual experiment in order to investigate the patterns of emotional speech recognition was carried out.

Thirty Russian adults were asked to listen to the files selected out of the Aibo Emotion Corpus. All the listeners were native speakers of Russian with no knowledge of German. The stimuli were short phrases that had been pronounced by German children in situations which evoked emotional verbal reactions. The subjects of the experiment had to make a decision which emotion was expressed in each phrase. They were asked to select from the list of 11 types of emotional states. The same list of emotions had been used by German annotators of the corpus (Table 1) [9]. The Table 1 shows the frequency of state evaluations done by Russian (for the experiment material) and German listeners (the whole Aibo corpus).

The comparison of evaluations showed a quite interesting tendency. In some cases a phrase that had been labeled as neutral (N) by the German annotators was consistently evaluated as surprised (S) by the Russian listeners. However, in most cases the Russian evaluations were identical to the German ones. This confirms that certain types of emotions have common means of expression in terms of acoustics in both languages while others differ. Thus the Russian corpus of emotion-related states is supposed to include samples of speech produced in different functional and emotional states which expression is specific for Russian. The acoustic characteristics of the emotions having common forms of expression can be obtained using the German data.

Table1. Frequency of State Evaluations: Russian vs. German

| | **Russian (111 files)** | | **German (the whole corpus)** | |
|---|---|---|---|---|
| State | abs. freq. | rel. freq. | abs. freq. | rel. freq. |
| neutral (N) | 611 | 18.5% | 172,575 | 71,30% |
| emphatic (E) | 501 | 15.2% | 39,942 | 16,50% |
| bored (B) | 369 | 11.2% | 10,413 | 4,30% |
| surprised (S) | 343 | 10.4% | 4,416 | 3,10% |
| touchy (T) | 265 | 8% | 5,924 | 2,40% |
| hesistant (H) | 294 | 8.9% | 1,594 | 0,60% |
| motherese (M) | 229 | 6.9% | 1,487 | 0,60% |
| joyful (J) | 214 | 6.5% | 1,153 | 0,50% |
| reprimanding (R) | 202 | 6.1% | 880 | 0,40% |
| angry (A) | 159 | 4.8% | 584 | 0,20% |
| other (O) | 113 | 3.4% | 37 | 0,00% |
| **Total** | **3300** | | **1701,5** | |

## 3. PERCEPTION OF RUSSIAN NON-FINAL INTONATION BY FINNISH SPEAKERS

Some contours in languages are similar both in their generalized form and function. Thus, rising intonation is thought be a universal feature of interrogation and non-finality [3]. At the same time concrete patterns may be language specific. Russian rising-falling intonation, the most common for general questions and non-finality have always been considered as such. Observation over language behavior of foreign speakers of Russian allows us to admit that this pitch pattern is often misinterpreted phonetically, phonologically and, consequently, pragmatically.

Unlike Russian, Finnish has a restricted range of intonation choice and has often been described as monotonous [5]. With falling intonation both for statements and questions, the distinctive feature remains the initial pitch of the utterance: in questions it is higher [11: 168-169].

Whether or not there are rising pitch contours in Finnish has been a matter of long dispute. Rising intonation contours used to be considered unacceptable due to the internalized prosodic grammar of native speakers of Finnish [5: 43]. In general Finnish still continues to be regarded a "level" rather "contour" language: for incomplete utterances Finnish speakers tend to use a mid- level tone, the non-low pitch throughout the intonation-group, with the first and the last on the same level [4: 27-29], But under certain specific circumstances, rising contours do occur in the intonation system of Finnish. Iivonen regarded the possibility for continuation rising intonation at the end of an intonation group, and a steeply ascending rise for echo questions [5: 51]. [11: 170] report that "rising" tone serves as a prosodic politeness formula. Rising intonation is thought to be characteristic of the speech of people living in Northern Finland. At the same time there is some evidence that rising

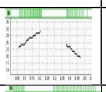
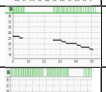
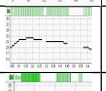
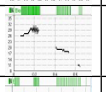
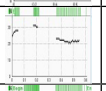
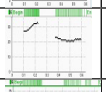
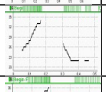
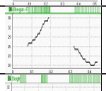
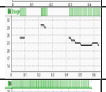
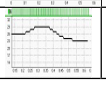
intonation contours are becoming more and more popular in colloquial Finnish with female speakers in Helsinki area [7]. Along with typical level continuation intonation when the pitch level at the end of the non-final utterance remains around the middle of the F0 range, without a fall, a slight rise may also occur [10: 116]. In objective descriptions Finnish, it is now widely recognized that intonation (e.g. rising intonation) serves quite specific functions in Finnish though these distinctions are used less systematically in Finnish than in some other languages [14], [13]. Whether or not rising intonation in Finnish is "latent", how Russian rising intonation contours "transplanted" into Finnish will be interpreted by Finnish speakers, whether or not they lead to misinterpretation — these issues we tried to elicit by conducting the perception experiment.

For this experiment sentences containing from 2 to 4 IPs were translated from Russian into Finnish and checked for accuracy by a native Finnish speaker. The experimental material was recorded from 4 Russian students of Finnish; their Finnish competence level can be estimated as pre-intermediate, and their pronunciation as acceptable with a strong Russian prosodic accent. The sentences were composed in such a way as to induce the speakers to use one of the Russian non-final contours: low rising, high rising or rising-falling. Recordings were performed in a sound –proof chamber at the Department of Phonetics of Saint-Petersburg State University.

Perception experiment was performed at the language lab of the Jyvaskyla University, in two sessions with 15 native Finnish speakers participating (Thanks to all the Finnish students for their participation). 44 stimuli containing intonation units with three different types of non-final intonation were selected from the material recorded from 4 Russian speakers of Finnish. The subjects were requested to read the instruction, listen to the stimuli presented one after another and tick the box, corresponding to their choice of the answers: Tone (**R**ising / **F**alling); Sentence type (**Q**uestion-**S**tatement(final)-**Exc**lamation-**O**ther (non-final); **N**eutral / **Em**otional. The stimuli were presented via headphones, the subjects were not limited in time. The results of the perception experiment are presented in Table 2.

First of all, the stimuli 5, 8, 22, 23, 34, 37 perceived as falling by Finnish subjects represent another type of rising intonation from the point of view of Russian speakers, a very slight lowering of the F0 in the post-tonic part is generally ignored. The contour itself sounds more familiar to the listeners as the one which is used in Finnish statements: due to this particular lowering of the post-tonic. This explains the number of "statement" judgments. The data also show a clear mismatch between the intonation of non-finality in Russian and Finnish both in the realization and perception of rising-falling intonation. Patterns 24, 27, 40, 17, 18 were judged as falling by the majority of the subjects. At the same time, there is no agreement between them as regards to the sentence type. They are no more unanimous in their "neutral /emotional" judgments.

Table 2. Perception of rising types of Russian non-final intonation by 14 Finnish listeners (%). Refusals are excluded.

| N | F0 track | Fall, % | Sentence type, % | | | | N/E, % | |
|---|---|---|---|---|---|---|---|---|
| | | | Q | S | Exc | O | N | Em |
| 24 |  | **79** | 21 | 21 | 43 | 14 | 43 | 57 |
| 27 |  | **79** | 0 | 50 | 21 | 14 | **86** | 14 |
| 22 |  | **93** | 0 | **79** | 14 | 7 | **86** | 21 |
| 23 |  | **79** | 0 | **93** | 0 | 7 | **86** | 14 |
| 40 |  | **100** | 28 | **64** | 4 | 4 | **86** | 14 |
| 5 |  | **86** | 0 | **79** | 7 | 7 | **79** | 21 |
| 8 |  | **86** | 50 | 7 | 43 | 0 | 7 | **93** |
| 17 |  | **71** | 21 | 21 | 36 | 14 | 43 | 50 |
| 18 |  | **71** | 28 | 28 | 28 | 14 | 57 | 36 |
| 37 |  | **71** | 0 | **71** | 21 | 7 | **79** | 21 |
| 34 |  | **71** | 7 | **57** | 0 | 36 | **100** | 0 |

So far we found no confirmation of the Toivanen's statements about the connection between the rising-falling shape of the contour and its affect-signaling property for Finnish. "With the respect to the affect-signaling property of the shape of the pitch-pattern, it has been established that in Finnish intonation a rising contour indicates that the speaker is uncertain. In contrast, by using a rising –falling (…) pitch contour, the speaker can emphasize a point very forcibly. Thus rise-fall often marks off a contrastive or emphatic focus. On the other hand, it seems that an abrupt pitch movement, whether downward or upward, reflects emotional arousal." [11: 172]. It seems that rising-falling intonation in our perception experiment is not interpreted as such since there in no distinctive phonological opposition between this type of contour and the rest: the rise-fall in

Finnish is used only as the accentual tune. A larger experiment is required. Anyway, the data show that functionally the same meaning — non-finality in our case — in Russian and Finnish displays similarities and differences in its formal representation and consequently, pragmatic interpretation, which may lead to misinterpretation in real speech situation. This information may become a point of special interest in the studies of L1-L2 contact and prosodic interference in general [12, 13].

## 4. CONCLUSION

The results of the experiments show that for any pair of the languages compared we can define similar and different forms of the acoustic manifestation of various functional states and means of conveying intonational meanings. These data can be obtained using the existing speech corpora of one of the languages. The results of the experiments provide information on, first, what material should be collected to be representative of the language specific features; second, the results of the perception experiments confirm the effectiveness of the use of "meaningless" for the listeners research material for obtaining information about universal and language specific means of conveying emotional and communicative information, and third, these data are of pragmatic value since they allow to predict situations which may lead to miscommunication in L1-L2 contact. Besides, intelligent use of the "foreign" language speech corpus may yield important valuable information about one's native language system.

## 5. REFERENCES

[1] A. Batliner, S. Steidl, E. Noth, "Releasing a thoroughly annotated and processed spontaneous emotional database: the FAU Aibo Emotion Corpus". In: *Proc. Satellite Workshop of LREC 2008 on Corpora for Research on Emotion and Affect,* Marrakesh, Morocco, pp. 28-31, 2008.

[2] L. Besacier, E. Barnard, A. Karpov, T. Schultz, "Automatic speech recognition for under-resourced languages: A survey". *Speech Communication*, Vol. 56, pp. 85–100, 2014.

[3] D. Hirst, A. Di Cristo (Eds.), "Intonation Systems: A Survey of Twenty Languages", Cambridge University Press, 1994.

[4] A.P. Hirvonen, *Finnish and English Communicative Intonation*, Publications of the Department of Phonetics 8, University of Turku. 1970.

[5] A. Iivonen, "Is there interrogative intonation in Finnish?", In: *Proc. The first conference of the Nordic Prosody*, Lund, Sweden, pp. 43-53, 1978.

[6] M. Maxwell, B. Hughes, "Frontiers in Linguistic Annotation for Lower-Density Languages", In: *Proc. COLING/ACL2006 workshop "Frontiers in Linguistically Annotated Corpora"*, Sydney, Australia, pp. 29–37, 2006.

[7] S. Routarinne, "Utterance Final Rises in the Speech of Adolescents from a Variational Perspective." In: *Proc. Hailuoto Workshop: New Trends in Variationist Linguistics*. 1998.

[8] R. Skadina I., A. Ahmet, V. Gioulli, D. Tufis, Gaizauskas, M. Mierina, N. Mastropavlos, "A Collection of Comparable Corpora for Under-resourced Languages". In: *Proc. the Fourth International Conference Baltic HLT 2010*, pp. 161-168, 2010.

[9] S. Steidl, *Automatic Classification of Emotion-Related User States in Spontaneous Children's Speech*, Erlangen, 2009.

[10] K. Suomi, J.Toivanen, R. Ylitalo, *Finnish Sound Structure. Phonetics, phonology, phonotactics and prosody*, Studia humaniora ouluensia 9, Oulu, 2008.

[11] J. Toivanen, "Perspectives on Intonation: English, Finnish and English Spoken by Finns". In: *Proc. Forum Linguisticum*, Band 3. Peter Lang, Frankfurt am Main, 2001.

[12] R. Ullakonoja, "Pitch contours in Russian yes/no questions by Finns", In: *Proc. Speech Prosody 2010 Conference*, Chicago, USA, 2010.

[13] R. Ullakonoja, "How Do Native Speakers of Russian Evaluate Yes/no Questions Produced by Finnish L2 Learners?", *Rice Working Papers in Linguistics*, Vol .5, pp. 92-105, 2010.

[14] R. Välimaa-Blum, "Intonation: A Distinctive Parameter in Grammatical Constructions". *Phonetica*, Vol. 50, pp. 124-137. 1993.