D.O.I. http://dx.doi.org/10.1590/2176-4573p55822

ARTIGOS

Os princípios da Escola Fonológica de São Petersburgo para a elaboração de *corpora* de fala / *Principles of the St. Petersburg Phonological School in Speech* Corpora *Design*

Pavel Skrelin[*]
Tatiana Kachkovskaia[**]
Daniil Kocharov[***]
Vera Evdokimova[****]
Uliana Kochetkova[*****]

RESUMO
O artigo discute os princípios fundamentais de elaboração do projeto e anotação de *corpora* de fala no âmbito da Escola Fonológica de São Petersburgo e fornece os exemplos de utilização de dados de vários *corpora* na pesquisa em fonética. Um dos princípios fundamentais é analisar as amostras em todos os níveis: desde o segmento até a entoação, incluindo as disfluências da fala. Durante a anotação fonética, sugerimos ouvir cada som isoladamente e confiar nos dados do espectrograma. Na anotação silábica, é crucial considerar a ressilabificação. Durante a anotação prosódica, sugerimos confiar na percepção do ouvinte e analisar as curvas melódicas. Um *corpus* de fala que segue esses princípios é uma fonte valiosa de dados fonéticos, uma vez que os fatores segmentais e prosódicos estão em constante interação e não se pode analisar as unidades de um nível de anotação sem fazer referência aos outros.
PALAVRAS-CHAVE: Fonética; Fonologia; *Corpus* de fala; Anotação fonética; Escola Fonológica de São Petersburgo

ABSTRACT
The paper discusses the main principles in designing and annotating speech corpora *within the framework of the Saint Petersburg phonological school, and provides examples of using* corpus *data in phonetic research. One of the major principles that we follow is to analyse the speech material at all levels: from segmental to intonational, including speech disfluencies. During segmental phonetic annotation, we suggest listening to each speech sound in isolation (without knowing its context) and relying on spectrographic data. At the syllabic tier, it is crucial to reflect resyllabification. During prosodic annotation, we suggest to rely on listener's perception of the intonation pattern first, then analyse the actual melodic curves. A speech* corpus *with multi-level annotation that*

[*] Saint Petersburg State University, Department of Phonetics, São Petersburgo, Rússia; https://orcid.org/0000-0002-8355-7378; skrelin@phonetics.pu.ru
[**] A produção do manuscrito ocorreu enquanto a autora trabalhava na Saint Petersburg State University, Department of Phonetics, São Petersburgo, Rússia; https://orcid.org/0000-0002-8588-9165; kachkovskaia@phonetics.pu.ru
[***] A produção do manuscrito ocorreu enquanto o autor trabalhava na Saint Petersburg State University – Department of Phonetics, São Petersburgo, Rússia; https://orcid.org/0000-0002-7858-5331; kocharov@phonetics.pu.ru
[****] Saint Petersburg State University, Department of Phonetics, São Petersburgo, Rússia; https://orcid.org/0000-0001-9742-5299; postmaster@phonetics.pu.ru
[*****] Saint Petersburg State University, Department of Phonetics, São Petersburgo, Rússia; https://orcid.org/0000-0003-1792-6064; u.kochetkova@spbu.ru

follows these principles is a valuable source of phonetic data — as segmental and prosodic factors are in constant interaction with each other, and one cannot analyse units of one annotation tier without reference to other tiers.
KEYWORDS: Phonetics; Phonology; Speech corpus*; Speech annotation; St. Petersburg phonological school*

Introdução

Atualmente, grande parte da pesquisa fonética é baseada em dados de *corpora* (LIBERMAN, 2019). Os primeiros grandes *corpora* de fala continham anotação detalhada feita manualmente. Por exemplo, o banco de amostras TIMIT *Acoustic-Phonetic Continuous Speech Corpus* [Banco Acústico-Fonético de Fala Contínua], que inclui as gravações de frases lidas por 630 falantes de inglês de várias partes dos EUA, é meticulosamente segmentado em sons de fala (GAROFOLO et al., 1993). Outro banco de amostras de fala bem conhecido, BURNC, *Boston University Radio News Corpus*, é notável devido à sua anotação prosódica feita de acordo com o sistema ToBI, *Tone and Break Indices* (OSTENDORF et al., 1995).

Com o passar do tempo, à medida que os *corpora* cresciam em tamanho, tornou-se óbvio que a anotação manual consumia muito tempo. O desenvolvimento de ferramentas de processamento de fala, no entanto, permitiu que os pesquisadores aplicassem os procedimentos automáticos de anotação e segmentação baseados em ASR (*Automatic Speech Recognition*). Com *corpora* de fala extremamente grandes, como *Librispeech* (PANAYOTOV et al., 2015) e *VoxPopuli* (WANG et el., 2021), a anotação manual tornou-se praticamente impossível. O *VoxPopuli* contém 400.000 horas de gravações de discursos pronunciados no Parlamento Europeu em 23 línguas, das quais são transcritas, mas não anotadas, apenas 1.800 horas. Apesar da ausência da anotação, esses *corpora* de fala são adequados para a pesquisa em fonética (ver, por exemplo, CHODROFF; WILSON, 2017).

Essa tendência recente, no entanto, não diminui de forma alguma a importância dos *corpora* de fala anotados para a pesquisa em fonética. Bancos como TIMIT e BURNS ainda apresentam o padrão de excelência na área da pesquisa de fenômenos segmentais e suprassegmentais em inglês.

A maioria dos *corpora* de fala são acompanhados por manuais de anotação que costumam descrever a terminologia e as regras de anotação. Os princípios de anotação

discutidos eventualmente levantam muitas questões na mente dos leitores atentos. Este artigo tem como objetivo discutir os princípios da elaboração e anotação de *corpora* de fala adotados no Departamento de Fonética da Universidade Estatal de São Petersburgo.

O artigo começa com um breve esboço histórico que descreve os primeiros *corpora* de fala armazenados no departamento. Em seguida, avança com uma discussão detalhada dos princípios de segmentação e anotação em cada nível particular – desde o fonético até o entoacional e paralinguístico. Finalmente, fornecemos alguns breves resumos sobre os *corpora* de fala mais notáveis desenvolvidos no departamento; esta seção termina com os exemplos de pesquisa fonética realizada com base nesses bancos.

1 Contexto histórico

O Departamento (‘Gabinete’) de Fonética Experimental em São Petersburgo foi fundado em 1899. Durante mais de cem anos, a pesquisa fonética aqui usou as mesmas ferramentas que os outros laboratórios similares no mundo: desde os diapasões e quimógrafos até os grandes *corpora de* fala e modelos de produção de fala. O atual Departamento de Fonética e Métodos de Ensino de Línguas Estrangeiras[1]da Universidade Estatal de São Petersburgo tem uma longa história de elaboração e adesão aos procedimentos para trabalhar com as gravações de *corpora* de fala extensos. Nas décadas de 1950 e 1980, a pesquisa fonética baseou-se em bancos de gravações de fala com as descrições detalhadas, quilômetros de filmes com oscilogramas e espectrogramas. O maior banco de amostras de fala elaborado na época dos anos 1970-1980 incluía as gravações em áudio produzidas por falantes de diferentes variantes regionais da Rússia. As gravações foram feitas em mais de 70 grandes cidades da União Soviética, pelo menos 20 falantes do sexo masculino de cada local participaram da pesquisa – predominantemente, os estudantes residentes na região ou na república onde acontecia a coleta de dados que falavam a variante regional (dialetal) do russo ou a variante local com alguns traços de sua primeira língua. O material que caracteriza, do ponto de vista fonético, cada variante dialetal do russo foi reunido em uma fita de áudio, e os resultados

[1] Daqui para a frente, Departamento de Fonética

desta pesquisa fonética foram publicados no final dos anos 1980 (BONDARKO; VERBITSKAYA, 1987).

No final da década de 1980, começou o desenvolvimento do Banco de Dados Automatizado da Língua Russa [*Mashinnyj fond russkogo iazyka*]. A coleção de dados fonéticos que representa um micromodelo digital para o sistema de sons do russo era destinada a ser parte considerável desse banco de dados (BONDARKO et al., 1992). O progresso atingido no desenvolvimento da Coleção Fonética do Russo [*Foneticheskii fond russkogo iazyka*] e os resultados da pesquisa relacionada tinham sido publicados escrupulosamente na revista especializada *The Bulletin of the Russian Phonetic Collection* [*B'ulleten' Foneticheskogo Fonda Russkogo Iazyka*; ou *Boletim do Banco de Dados Fonético da Língua Russa*]. A revista era publicada desde 1988 em Bochum, na Alemanha, e mais tarde parcialmente em São Petersburgo, na Rússia. Havia dois editores: o Prof. Christian Sappok, da Universidade de Ruhr, Bochum, e a Prof. Lia V. Bondarko, chefe do Departamento de Fonética da Universidade Estatal de Leningrado (mais tarde, Universidade Estatal de São Petersburgo), Rússia. O desenvolvimento do Banco de Dados Automatizado do Russo foi interrompido no início dos anos 1990. No entanto, nessa altura, a criação da Coleção Fonética do Russo estava quase concluída e, em 1993, o grupo de pesquisa publicou o Apêndice 3 ao Boletim do Banco Fonético do Russo intitulado 'As coleções de unidades sonoras da fala em russo' (Apêndice 3, 1993).

Os apêndices do Boletim sempre incluíam uma fita de áudio com o material de pesquisa, que foi posteriormente substituída por um CD. O apêndice 3 do Boletim do Banco Fonético do Russo foi publicado junto ao artigo de Bondarko (1993) sobre o sistema de sona do russo acompanhado por uma descrição pormenorizada e ilustrada de todos os componentes (módulos) do banco de dados. A seção 'A sílaba' incluiu espectrogramas e gravações de todas as 186 sílabas russas com a estrutura CV (consoante vogal) e V (vogal), com a segmentação em sons. A seção 'A palavra' contém 150 palavras transcritas com a ortografia opaca, 250 palavras do *Basic Learner's Dictionary* (PAPERNO; LEED, 1988) e do dicionário de frequências de russo. A seção 'O texto' inclui um texto foneticamente representativo, isto é, um texto de duas páginas que contém os fonemas e sílabas do russo mais frequentes e um diálogo relativo que ilustra o sistema de entonação russo. Todos os materiais de áudio, exceto os dados do dicionário de frequência, foram gravados por quatro falantes (2 homens, 2 mulheres), representando

duas variantes principais da pronúncia russa padrão: de Moscou e de Leningrado (São Petersburgo).

2 Os princípios da elaboração e anotação de *corpus* de fala

As ideias que surgiram durante o trabalho com o Banco Fonético do Russo e a experiência obtida nesses anos permitiram formar uma base para a elaboração de bancos de dados fonéticos para outras línguas faladas na Rússia. Essa tarefa exigia uma revisão dos princípios: da coleta e organização de dados; avaliação da qualidade através das experiências auditivas ou do equipamento técnico; escolha do *software* e equipamentos. Os princípios fundamentais adotados para a elaboração do Banco Fonético do Russo foram ajustados de acordo com as características específicas das línguas a descrever. Afinal, o banco consolidou a seguinte estrutura:

1. os dados em formato de áudio;
2. as características fonéticas de cada uma das unidades mínimas significativas da língua (fonemas/sílabas);
3. as estruturas fonéticas das formas de palavras;
4. a conversão automática de grafema para fonema;
5. as características fonéticas das unidades entoacionais.

O *corpus* de fala anotado pressupõe a anotação das gravações feita com a segmentação e marcação das unidades da fala nos níveis segmental e suprassegmental. Mais adiante, discutimos os princípios da anotação e marcação para um banco de amostras de fala.

O princípio da segmentação. Na anotação de um banco de amostras de fala, seguimos o princípio da divisão estrita em camadas: cada unidade de uma camada inferior deve ser incorporada por inteiro em uma única unidade de uma camada superior. Como resultado, não permitimos que os níveis de fronteira de fonema mais altos se situem dentro de um fonema. Este princípio facilita o processamento automático, embora exija uma série de regras de segmentação complementares.

Na fala contínua, muitas vezes observamos as omissões, inserções ou fusões de fonemas, que exigem uma atenção especial. Em caso de omissões, o fonema ausente

muitas vezes deixa algum traço: por exemplo, uma vogal labializada, quando não pronunciada, ainda causa o arredondamento das consoantes precedentes (por exemplo, em russo *существование*, em português "existência": [sʷʃʲːistvʌˈvanʲi̯i]).[2] No caso das inserções, por exemplo, das vogais em alguns *clusters* consonantais, precisamos decidir o que conta como um fonema, porque as inserções são frequentemente muito curtas (por exemplo, em russo *корабль*, em português "navio": [kʌˈrabəlʲ]).[3] As fusões, por definição, são difíceis de serem identificadas. Quando os dois sons idênticos ocorrem um após o outro, não há limite claro entre eles. Nesse caso, podemos marcar a fronteira de fonema diretamente no meio do som, mas esse método causa polêmica quando lidamos com as paradas geminadas, pois em tais casos as explosivas são produzidas apenas uma vez (por exemplo, em russo *оттуда*, em português "daí": [ʌˈt̄ tudʌ]). Estas questões podem ser parcialmente resolvidas através da anotação da fala em diferentes níveis segmentais. Isso permite descrever um fonema como constituído de vários sons e um único som correspondente a vários fonemas.

Uma única e evidente exceção para o princípio da segmentação é observada no nível silábico devido ao fenômeno da ressilabificação (por exemplo, em russo *брат Ани*, em português "irmão de Anya": [ˈbrat ˈanʲi], [bra-ta-nʲi]). Para algumas línguas, a divisão em pés rítmicos também pode desafiar este princípio (por exemplo, em inglês *come again*, em português "volte sempre": [ˈkʌmə-ˈgen]).

Definição das fronteiras. As fronteiras dos segmentos devem ser detectadas com a maior precisão possível. Isso garante a precisão alta das fronteiras nas camadas de anotação mais altas. Há uma série de recomendações publicadas para a segmentação, por exemplo (TURK et al., 2012). Os princípios de segmentação podem basear-se nas tarefas específicas para as quais o banco de amostras de fala é constituído. Em geral, os rótulos são colocados nas fronteiras das realizações físicas dos alofones. A segmentação deve satisfazer o seguinte critério: os alofones finais podem ser transplantados para as outras palavras com o mesmo tipo de som (SKRELIN, 1999).

[2] Curiosamente, quando uma vogal é omitida, o número percebido de sílabas não se altera: neste exemplo, a sequência [sʷ] ainda forma uma sílaba por si só. Assim, a estrutura rítmica da palavra permanece intacta.
[3] Comumente, essas inserções vocálicas não alteram o número de sílabas percebidas. Ou seja, semelhante ao exemplo anterior, a estrutura rítmica da palavra permanece intacta.

Duas camadas para a anotação fonética. O nível fonético ‘acústico’ é o nível fonético comumente aceito que pode ser encontrado em muitos *corpora de* fala conhecidos. É feito por meio da escuta dos sons analisados em contexto isolado e da utilização dos métodos instrumentais (espectrogramas). Os princípios que constituem este nível são: (1) a deslexicalização e (2) a consideração das propriedades acústicas dos sons para garantir a máxima precisão e objetividade da transcrição.

A deslexicalização visa resolver os problemas das interpretações fonéticas causadas pelo ouvido do foneticista. A ausência deste princípio leva a decisão do anotador ser influenciada pelo seu próprio conhecimento do conteúdo fonêmico da palavra em questão (BONDARENKO et al., 1974). As decisões do anotador devem se basear em dados precisos, nas propriedades físicas dos sons. Isso é especialmente crucial para as vogais, no caso das quais devemos confiar nos valores dos formantes (para os dados sobre as vogais do russo, ver EVDOKIMOVA et al., 2020).

A transcrição fonética ‘perceptiva’ é realizada através da escuta da palavra como um todo. Essa transcrição permite revelar as particularidades percebidas da pronúncia do falante, incluindo os traços específicos de uma determinada região. Este tipo de transcrição diferencia-se da transcrição fonética acústica porque se baseia no conhecimento do anotador sobre o padrão de pronúncia (ou dialeto principal). Como resultado, esta camada contém apenas informação relevante para a percepção, que é facilmente detectada pelo ouvido (por exemplo, em russo *водяной*, em português “espírito da água”: [vʌdʲaˈnoi] em vez de [vʌdʲiˈnoi], o que é típico para algumas regiões da Rússia). Para obter a transcrição adequada, os anotadores devem ter conhecimentos similares sobre a pronúncia padrão.

Anotação fonêmica. A transcrição fonêmica é baseada em regras ortoépicas, conforme descrito nos dicionários de pronúncia. No entanto, esses dicionários contêm palavras soltas; na fala contínua, o conteúdo fonêmico da palavra, muitas vezes, sofre alterações por conta dos processos de assimilação (compare, por exemplo, em russo *под березой*, em português “sob a bétula”, e em russo *под пихтой*, em português “sob o abeto”: /pad-biˈrʲozaij/ e /pat-ˈpʲixtaj/). Quando há boa descrição dessas alterações fonêmicas para a língua, a transcrição fonêmica automática pode apresentar uma boa qualidade; no entanto, o *software* da transcrição automática exigirá as informações sobre as fronteiras prosódicas

e pausas, porque os processos de assimilação comumente não ultrapassam as fronteiras prosódicas grandes.

Lidando com os processos de assimilação, podem-se enfrentar as outras dificuldades na transcrição da fala contínua. Em alguns casos, observamos os sons ausentes no sistema fonológico da língua. Por exemplo, no sistema fonológico do russo, o traço 'vozeado-desvozeado' está presente na maioria das articulações: /p-b/, /t-d/, /s-z/ etc. No entanto, alguns fonemas não têm as suas contrapartes vozeadas ou desvozeadas, por exemplo, o fonema / ʃʲː/. Devido à assimilação regressiva, alguns contextos podem levar a variante vozeada /ʒʲː/ (e.g. em russo *плащ дедушки*, em português *o casaco do avô*: [ˈplaʒʲː ˈdʲeduʃkʲi]). Como resultado, mesmo as regras de pronúncia mais básicas acabam não sendo descritas em termos de fonemas; basicamente trabalhamos com os alofones e as transcrições resultantes são alofônicas.

Vale a pena ressaltar, no entanto, que esse tipo de transcrição baseada em regras ainda é fonológico, não fonético. O número possível de rótulos para os alofones é apenas um pouco maior do que o número de fonemas. O nível puramente fonêmico pode ser adicionado, caso necessário, e pode ser facilmente gerado automaticamente.

Nível silábico. Em línguas diferentes e dentro de diferentes tradições linguísticas, as fronteiras silábicas são definidas de diferentes maneiras. Entre as abordagens mais comuns estão o princípio da sonoridade e o princípio da distribuição. Na tradição da escola fonológica de São Petersburgo, um outro princípio é usado: a fala em russo é dividida em sílabas abertas. Este princípio foi formulado após uma série de experimentos de produção de fala com '*feedback* atrasado' (também chamado de 'gagueira artificial') realizados por Chistovich e Bondarko (1963). No experimento, os pesquisadores colocaram um palato artificial na boca dos participantes que os impedia de sentir sua própria articulação; ao mesmo tempo, os participantes usavam fones de ouvido nos quais as gravações da sua própria fala estavam tocando com um atraso significativo. Descobriu-se que os falantes nunca produzem uma pausa dentro das sequências CV (consoante vogal), enquanto as consoantes em coda podem ser separadas e formar uma nova sílaba, muitas vezes com a adição de uma vogal neutra.

Com tempo, o princípio da sílaba aberta adquiriu uma série de exceções. Entre as mais notáveis estão as sílabas que terminam em consoante vocalizada: por exemplo, em

russo *майка*, em português "regata": [ˈmaɪ̯-kʌ]. Neste caso, a divisão em sílabas abertas teria produzido a sílaba [ɪ̯kʌ] que, se reproduzida por si só, é percebida como duas sílabas em vez de uma. Outra exceção refere-se às consoantes em coda nas extremidades de grandes unidades prosódicas: esses sons podem ser considerados quase sílabas. Caso necessário, as fronteiras silábicas podem ser automaticamente alteradas em relação a outros princípios de silabificação.

Nível de palavra. Em línguas como o russo, a forma textual de uma frase não é facilmente combinada com a pronúncia real, especialmente no que diz respeito à colocação de acento. Por exemplo, uma frase preposicional costuma ser pronunciada com apenas um acento, mas não necessariamente; isso depende da própria proposição, da estrutura lógica da frase e de outros fatores. Como resultado, a anotação é comumente realizada em dois níveis diferentes: no nível da transcrição ortográfica, em que há o espaçamento entre as unidades lexicais; e no nível da transcrição fonética, em que uma ou mais palavras lexicais são unidas por um único acento (por exemplo, em russo *не де́лали бы*, em português "não fariam"). Entre outras razões, o nível das unidades fonéticas é crucial para a pesquisa em entoação, pois muitos fenômenos entoacionais são ancorados nas sílabas tônicas.

Se a língua permite os acentos secundários, são necessárias várias regras de segmentação complementares. Uma solução possível é marcar o acento principal que forma as palavras fonéticas, bem como alguns graus adicionais de acento mais fraco. Não apenas as palavras compostas apresentam as dificuldades, mas também os pronomes, as conjunções e outras palavras que são flexíveis em termos de colocação de acento. Um exemplo curioso do russo é a pronúncia da conjunção *но* (em português, "mas"), que muitas vezes funciona como um proclítico, mas sempre preserva a sua qualidade vocálica ([no]) – apesar do fato de que em russo /o/ em posições átonas é reduzido a /a/. Na prática, nenhuma solução é perfeita, pois a adição de mais tipos de acentos dificulta a convenção entre os anotadores e reduz a precisão da transcrição automática.

Entoação. As descrições tradicionais da prosódia do russo são semelhantes às da Escola Britânica (como em O'CONNOR; ARNOLD, 1973). A unidade básica de segmentação é a frase entoacional (IP, *Intonational Phrase*), tal que:

- uma IP geralmente contém uma palavra principal (o núcleo) em torno da qual é realizada uma curva melódica linguisticamente relevante;
- uma IP é percebida como um todo em termos de melodia, intensidade e posição de pausas;
- uma IP não pode ter mais de um núcleo (exceto para os padrões melódicos específicos em que dois núcleos são obrigatórios);
- certos fenômenos prosódicos ocorrem nos limites de IP (por exemplo, alongamento pré-fronteira).

Esta definição é altamente adequada para a fala planejada e requer mais especificações para a fala com disfluências, em que as IPs são frequentemente inacabadas ou contêm um elemento paralinguístico interno, por exemplo, um preenchimento ou uma pausa silenciosa. Como resultado, temos de admitir que uma IP não necessariamente deve conter um núcleo (IPs inacabadas), enquanto os elementos paralinguísticos internos à IP podem induzir alguns fenômenos de fronteira no meio de uma IP.

A anotação prosódica deve ser realizada por foneticistas profissionais com base na análise auditiva e instrumental. Geralmente, esse trabalho requer formação prévia do profissional para garantir a convenção entre os anotadores. Na maioria dos casos, os rótulos prosódicos são adicionados à transcrição ortográfica. Em todos os *corpora* de fala prosodicamente anotados, a anotação inclui as seguintes informações prosódicas: as fronteiras das frases entoacionais (IPs); a localização da palavra principal dentro da IP (para as frases que contêm núcleo); o tipo de movimento melódico da IP; as palavras com a proeminência prosódica complementar. Além do núcleo, alguma outra sílaba pode ter proeminência prosódica adicional, conforme percebido pelo anotador. Tal proeminência pode ser manifestada por qualquer tipo de parâmetros prosódicos – mudanças notáveis na frequência fundamental (F0), intensidade, duração, qualidade de voz ou por combinações desses fenômenos (VOLSKAYA; KACHKOVSKAIA, 2016).

Limites dos períodos de pitch. Para alguns fins, podemos precisar de uma descrição precisa do contorno melódico. Às vezes, os algoritmos de detecção automática de *pitch* cometem erros (por exemplo, erros de duplicação/redução pela metade) que resultam em cálculos errôneos dos principais parâmetros prosódicos. No entanto, para um foneticista experiente, esses erros são fáceis de serem percebidos e corrigidos, caso o *software*

proporcione essa opção. Isso pode ser realizado através das análises simultâneas: auditiva da gravação e visual da onda sonora. Para obter os valores de F0 com uma precisão maior, é necessário marcar manualmente as fronteiras dos intervalos de *pitch*; essa fase geralmente é precedida por uma marcação automática, que apresenta alguns erros.

Dependendo da tradição linguística, os intervalos de *pitch* podem ser marcados em relação a diferentes pontos iniciais (por exemplo, em picos de amplitude no *Praat*). A escola fonológica de São Petersburgo adota a visão segundo a qual cada intervalo começa no ponto em que a amplitude é igual a zero, onde os valores mudam do negativo para o positivo. Essa visão é motivada pelo fato de este ponto corresponder ao início do primeiro formante (SKRELIN, 1999). Em geral, a escolha do ponto inicial não parece ter muita influência nos valores F0 resultantes.

Os fragmentos de fala produzidos em *creaky voice* são outra fonte de erros de detecção de *pitch* devido à vibração irregular das pregas vocais e, como resultado, à duração altamente variável dos períodos de *pitch* adjacentes. Nesses casos, mesmo a segmentação manual não permitiria obter os valores de F0 precisos. Por causa disso, nesse nível, os fragmentos permanecem sem anotação. Caso necessário, a marcação de *creaky voice* pode estar em uma camada especial de anotação, juntamente a outras configurações fonéticas.

As disfluências e os fenômenos não relacionados à fala. Muitas vezes, na fala espontânea e, raramente, na fala planejada, o fluxo de palavras é interrompido por pausas silenciosas, pausas preenchidas, hesitações não fonêmicas alongadas, risos, tosses, cliques casuais não fonêmicos etc. Além disso, qualquer gravação pode conter os eventos não relacionados à fala causados pelo equipamento de gravação ou pelo *software* de pós-processamento. Para um alinhamento automático mais preciso da transcrição ao sinal da fala, todos estes fenômenos exigem uma marcação especial.[4] Essa marcação pode também ser útil para os pesquisadores interessados nesses fenômenos específicos.

A definição das fronteiras desses fragmentos é repleta de dificuldades. Alguns destes fenômenos podem ocorrer simultaneamente com a fala, por exemplo, risos, tosse ou ruídos técnicos. Isso significa que as suas fronteiras devem ser marcadas em um nível

[4] A nossa experiência mostra que o acréscimo das disfluências da fala para a transcrição ortográfica pode aumentar até 10% o número total de sons da fala (KACHKOVASKAIA et al., 2016).

de anotação separado, ou mesmo em vários níveis separados. O outro problema diz respeito ao preenchimento de pausas (como ‘ehm’ e ‘uhm’): quando esse elemento começa logo após um fonema vocálico, especialmente uma vogal aberta, é difícil detectar o limite entre o fonema vocálico e o preenchedor. No entanto, para solucionar a maioria dessas dificuldades, não é necessária uma marcação precisa dos fenômenos desse tipo. É por isso que a melhor decisão é anotá-los em um dos níveis existentes, por exemplo, no nível da entoação. Alguns destes fenômenos podem também ser marcados automaticamente, em especial, as pausas silenciosas.

3 *Corpora de* fala anotados desenvolvidos no Departamento de Fonética

3.1 O *corpus* da fala em russo INTAS

Os princípios de anotação e segmentação apresentados neste artigo podem ser demonstrados da melhor maneira no exemplo do banco de amostras de fala em russo INTAS. Este banco foi criado dentro do projeto INTAS 915, o banco de amostras de fala espontânea de línguas tipologicamente não relacionadas - russo, finlandês e holandês (SKRELIN, 2009). Para o banco de amostras em russo, foram gravados dez falantes nativos da língua russa (5 homens, 5 mulheres). Os participantes representavam diferentes faixas etárias (abaixo de 20 anos, 20-30 anos, 30-40 anos, 40-50 anos, acima de 50 anos), todos os participantes eram falantes da variante padrão do russo, a de São Petersburgo.

A primeira etapa consistia na gravação de diálogos informais entre os participantes e seus familiares. De cada gravação, um fragmento de 5 minutos foi selecionado para a análise posterior. Na segunda etapa, cada participante foi convidado a ler um texto construído com base no seu próprio monólogo.

A segmentação foi realizada por diferentes foneticistas que seguiram as regras de segmentação usadas para concatenação de alofones no sistema de síntese de fala baseado em alofones (SKRELIN, 1999). Durante a segmentação, foi realizada a transcrição preliminar de segmentos. Ao mesmo tempo, o *pitch* foi automaticamente detectado e corrigido manualmente. A segmentação e a transcrição foram então verificadas e corrigidas por dois foneticistas experientes.

A estrutura de anotação contém 8 níveis (ver Fig. 1):

1. O nível acústico: a transcrição fonética (segmentação e marcação) produzida com base na escuta dos sons da fala isolados utilizando os métodos instrumentais (espectrogramas);
2. O nível perceptual: a transcrição fonética (segmentação e marcação) produzida com a base na escuta de sons dentro da palavra;
3. O nível fonêmico (o mais preciso): a transcrição fonêmica (segmentação e marcação) com base nas regras padrão de pronúncia da língua russa;
4. O nível silábico: a segmentação em sílabas abertas usando a transcrição do nível 1;
5. O nível do acento: a indicação do acento lexical;
6. O nível das palavras fonéticas: as palavras de conteúdo e os seus clíticos circundantes;
7. O nível das palavras ortográficas: as palavras separadas por espaçamento;
8. O nível das unidades de entoação: a segmentação e a marcação das frases entoacionais, com as informações sobre o tipo melódico para cada IP e tipo de pausa de acordo com o sistema de anotação sugerido por N. Volskaya (VOLSKAYA; KACHKOVSKAIA, 2016);
9. O nível da prosódia: a marcação dos principais movimentos melódicos (subida vs. descida);
10. O nível dos fenômenos da hesitação (pausas preenchidas): segmentação e marcação.

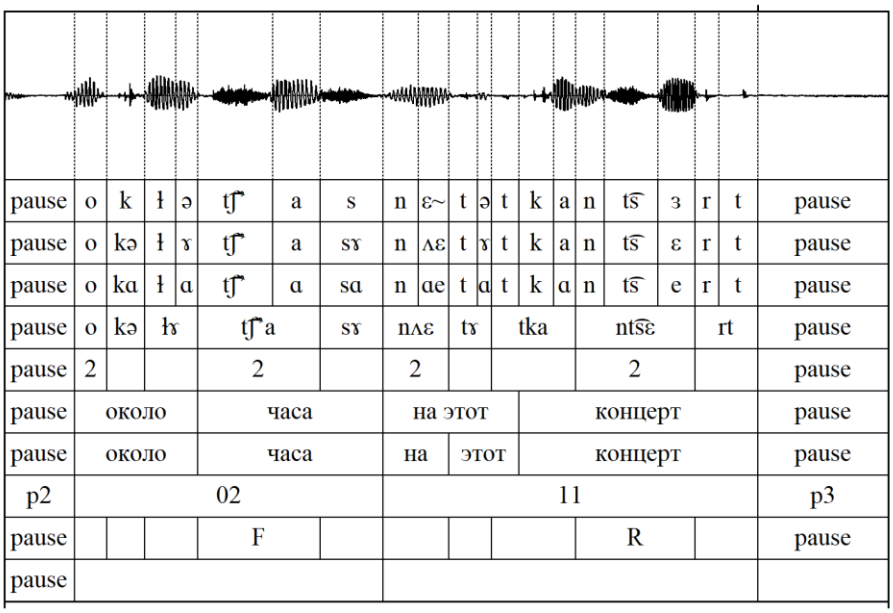

Figura 1. Os níveis de anotação no banco de amostras de fala INTAS

A Figura 1 mostra um exemplo de anotação da frase em russo *[времени было потрачено] около часа на этот концерт* (em português *[gastamos] por volta de uma hora para esse show*) realizada no *Praat*. Comparando os níveis acústico e perceptivo, podemos encontrar os casos de omissões vocálicas na pronúncia real. O nível silábico demonstra com clareza o princípio da sílaba aberta; há também um exemplo de ressilabificação (em russo *этот концерт*, em português, "esse show": sílaba [tka]) que desafia o princípio de divisão estrita silábica. No nível da entoação, podemos notar dois IPs[5] sem pausa entre elas e as pausas do lado esquerdo e do lado direito; o nível prosódico nos ajuda a ver onde exatamente os movimentos melódicos (a queda, marcada como 'F', e a ascensão, marcada como 'R') estão localizados. O nível da hesitação está vazio neste exemplo, mas em outras partes das gravações, ele marca as fronteiras das pausas preenchidas.

3.2 Outros *corpora de* fala

As informações sobre os outros bancos de dados mencionados neste artigo estão resumidas nas Tabelas 1 e 2. A seguir, apresentamos breves descrições dos *corpora* de fala mais notáveis desenvolvidos no Departamento de Fonética. Este resumo não inclui os bancos especializados de segmentos sonoros que foram desenvolvidos para os sistemas automáticos de síntese de fala em russo (SKRELIN, 1997a).

CORPRES, Corpus of Russian Professionally REad Speech [*Corpus* de fala lida por professionais] (SKRELIN et al., 2010) foi criado em 2009-2011 e originalmente destinado ao uso para a síntese da conversão do texto em fala baseado na seleção de unidades. No entanto, sendo uma boa representação da fala em russo padrão, tem sido usado em um grande número de projetos de pesquisa fonética. As gravações foram feitas por locutores profissionais[6], falantes da variante padrão de São Petersburgo.

CoRuSS, Corpus of Russian Spontaneous Speech [Banco de amostras de fala espontânea do russo] (KACHKOVSKAYA et al., 2016). O principal objetivo do trabalho foi criar um banco de amostras de fala não lidas gravadas por falantes de gêneros e faixas etárias

[5] Tipo melódico 02, isto é, o movimento tonal de (subida-)descida dentro do núcleo, frequentemente é usado para demonstrar a ênfase ou o contraste; tipo 11, isto é, p movimento tonal subida(-descida) frequentemente é usado em IPs não finais. O tipo de pausa p2 corresponde à uma quebra prosódica mais leve do que a do tipo p3.

[6] Principalmente as emissoras de TV e rádio.

diferentes. O *corpus* de fala destinava-se a ser utilizado para a pesquisa em marcação automática de fronteiras prosódicas. As gravações foram feitas em forma de diálogos espontâneos. Para além do diálogo espontâneo, cada participante gravou também o texto foneticamente representativo e um breve monólogo sobre si próprio.

Durante a elaboração desse banco de amostras, o sistema de anotação foi atualizado para incluir vários tipos de disfluências e fenômenos paralinguísticos. Abaixo, pode-se encontrar um fragmento do nível ortográfico contendo a anotação prosódica junto aos outros fenômenos variados. Neste exemplo[7], as barras correspondem às fronteiras de IP, [02], [09], [11] e [11b] representam os tipos melódicos (*tunes*), [+] representa a proeminência prosódica complementar, '9' corresponde ao riso, 'э-' é uma hesitação vocálica de qualquer qualidade, os rótulos '1' e '2' após as vogais são usados para marcar o acento forte e fraco, respectivamente, e os dois-pontos marcam os alongamentos dos sons da fala.

SibLing Corpus of Russian Dialogue Speech [*Corpus* de fala em forma de diálogo] (KACHKOVSKAYA et al., 2020). Esse banco de amostras foi desenvolvido especificamente para a pesquisa sobre o fala entrelaçada – o fenômeno de afinidade entre os falantes na conversação que resulta em semelhanças nos gestos, mímicas e fala dos interlocutores. No SibLing, a amostra básica de falantes consistia em 10 pares de irmãos do mesmo sexo com idade entre 23 e 40 anos. Cada um destes 20 falantes conversou com 5 interlocutores diferentes (participantes convidados) de diferentes graus de familiaridade e 'distanciamento social': de irmãos ou amigos próximos até os estranhos de idade significativamente maior. Durante a gravação, cada par de interlocutores realizou duas tarefas colaborativas: um jogo para encontrar os pares de cartas e uma tarefa de mapa.

Multimedia Corpus of Ironic Speech [*Corpus* do discurso irônico em formato multimídia] (KOCHETKOVA et al., 2021) foi desenvolvido no âmbito do projeto *Acoustic correlates of irony* [Correlatos acústicos da ironia], que diz respeito aos tipos básicos de movimento de *pitch*. O banco de amostras contém a leitura de 330 monólogos e diálogos breves, bem como quatro longos textos coerentes, que incluíam os enunciados homónimos irônicos e

[7] [02]не1 / ну [+]ла1дно та1м [02]преподава1тели / я2 ду1маю что2 они2 таки1е у на2с лю1ди [09]обеспе1ченные / 9 / а [11b]аспира1нтам / 9 / э- ну1 в осо1бенности ка1к [02]лёлха / 9 / и1м оставля1ют то1лько [11]с:типе1ндию / кото1рая [02]госуда1рственная / Em inglês: [02]no1 / well [+]oka1y tho1se [02]te1achers / i2 thi1nk tha2t the2y a2re ki1nd of pe1ople [09]be2tter-o1ff / 9 / and [11b]stu1dents / 9 / э- we1ll espe1cially li1ke [02]a1lex / 9 / the1y a2re le1ft with o1nly [11]s:cho1larship / whi1ch is [02]bu1dgetary /

não irônicos de vários tipos comunicativos, permitindo a implementação de todos os padrões melódicos possíveis. As conotações exigidas (irônico vs. não irônico) foram simuladas pelo contexto: por meio de marcadores lexicais, gramaticais ou semânticos de ironia, bem como apenas por contexto. Juntamente ao áudio, o banco de amostras contém as gravações de vídeo de alta velocidade.

Nome	Material	Participantes	Duração total das gravações	Acesso ao banco
INTAS Corpus of Russian Speech	5 minutos de fala espontânea + 5 minutos de leitura (o texto com o conteúdo lexical similar)	10 falantes, grupos de gênero e faixas etárias diferentes	1 h 40 min	Há acesso[8]
CORPRES	Textos de ficção e não-ficção lidos pelos locutores profissionais	4 homens, 4 mulheres	60 h	Não há acesso[9]
CoRuSS	Diálogos espontâneos (conversa livre)	60 falantes, grupos de gênero e faixas etárias diferentes	30 h	Há acesso
SibLing	Diálogos colaborativos (jogo de pares de cartas; tarefa de mapa)	100 falantes	64 h	Há acesso
Multimedia Corpus of Ironic Speech	Leituras de monólogos e diálogos curtos, textos coesos longos (inclui as sentenças homonímicas irônicas e não irônicas de vários estilos comunicativos).	56 falantes, grupos de gênero e faixas etárias diferentes	12 h	Há acesso

Quadro 1. *Corpora* de fala desenvolvidos recentemente no Departamento de Fonética da Universidade Estatal de São Petersburgo

Nível	***Corpus***				
	INTAS	**CORPRES**	**CoRuSS**	**SibLing**	**Ironic speech**
Acústico fonético	S(m), L(m)	S(m), L(m)			
Fonético perceptual	S(m), L(m)				
Fonêmico	S(m), L(m)	S(m), L(m)	L(a)	L(a)	
Silábico	S(m), L(m)				
Palavra	S(m), L(m)	S(m), L(m)	L(m)	L(m)	L(m)

[8] De hoje em diante: disponível para os fins académicos mediante solicitação (através do contato com os criadores).
[9] O banco de amostras pertence a uma empresa particular.

Acento	S(m), L(m)	S(m), L(m)	L(m)	L(m)	
Entoação	S(m), L(m)	S(m), L(m)	L(m)	L(m)	L(m)
Pausas	S(m), L(m)	S(m), L(m)	S(m), L(m)	S(m), L(m)	
Períodos de *pitch*	S(m), L(m)	S(m), L(m)		S(a), L(a)	
Disfluências	S(m), L(m)		S(m), L(m)	L(m)	
Fenômenos não linguísticos			S(m), L(m)	L(m)	

Quadro 2. A estrutura de anotação dos *corpora de* fala desenvolvidos recentemente no Departamento de Fonética da Universidade Estatal de São Petersburgo: S – segmentados, L – com marcação; a segmentação e a marcação são manuais (m) ou automáticas (a)

Os bancos e as coleções de amostras de fala. Existem também várias bancos e coleções de amostras de fala que foram anotadas parcialmente de acordo com os princípios descritos na seção 2:

- O banco de amostras de fala gravadas de V. M. Zhirmunsky (SVETOZAROVA, 1996);
- Os contos do Norte da Rússia (SKRELIN et al., 1997b);
- O folclore poético do Norte da Rússia (Lamentos) (SKRELIN, 1998);
- A fala em russo dos *doukhobor*[10] canadenses (MAKAROVA et al., 2011);
- As gravações de fala para a pesquisa em fadiga vocal (EVGRAFOVA et al., 2016);
- As gravações de cantores profissionais (EVDOKIMOVA et al., 2017).

4 *Corpora* de fala na pesquisa fonética

Muitos anos de experiência de coleta, processamento e análise de fala permitiram-nos criar *corpora* de fala de todos os tipos que podem servir de base para uma vasta gama de pesquisas fundamentais e aplicadas. O *corpus* de fala lida CORPRES serviu de base para muitos projetos de pesquisa, incluindo a pesquisa da marcação automática de fronteiras prosódicas (KOCHAROV et al., 2019a), a pesquisa sobre a redução vocálica (KOCHAROV et al., 2019b) e alongamento em final de frases (KACHKOVSKAIA et al, 2013), declínio da curva melódica (KOCHAROV et al., 2015), a melodia do pós-núcleo (KACHKOVSKAIA et al., 2020) e outros.

[10] N. do T.: *doukhobor* canadenses são dissidentes religiosos provindos da Rússia que emigraram do Império Russo para o Canadá no final do século XIX.

Os dois grandes *corpora* de fala (CORPRES e CoRuSS) incluem anotação que foi usada como material para comparar discurso lido e espontâneo em termos de duração de frase entoacional, distribuição de tipos melódicos, duração da pausa silenciosa, frequência de marcação de fronteiras de IP como pausas silenciosas (KACHKOVSKAIA; SKRELIN, 2020).

O *corpus* SibLing serve como a principal fonte de dados para a pesquisa sobre a fala entrelaçada (MENSHIKOVA et al., 2020). A organização específica deste *corpus* permite traçar a influência de fatores sociais e situacionais na fala dos interlocutores (KACHKOVSKAIA et al, 2022).

O *corpus* de fala irônica em formato multimídia serviu de base para a comparação das características fonéticas de fala irônica nas línguas russa e francesa (SKRELIN et al., 2021), percepção da ironia na fala masculina e feminina (Kochetkova et al., 2020).

A pesquisa sobre as pistas acústicas de fadiga vocal foi realizada com base nas gravações de falantes antes e depois de uso intensivo da voz (EVDOKIMOVA et al., 2017). A pesquisa sobre as pistas acústicas de patologias vocais na fala cantada foi realizada com base em gravações de cantores profissionais (EVDOKIMOVA et al., 2019).

Conclusão

Na pesquisa fonética, muitas vezes é crucial analisar a interação complexa entre os fatores que atuam nos níveis diferentes. É por isso que um dos princípios fundamentais que adotamos é analisar o material da fala em todos os níveis. Assim, um banco de amostras de fala anotado de maneira mais completa deveria incluir os níveis de anotação que refletem os princípios fundamentais do projeto do banco de amostras baseado em princípios da escola fonológica de São Petersburgo.

1) Nível fonético 1 (acústico). Os princípios fundamentais: (1) deslexicalização e (2) consideração das propriedades acústicas dos sons.
2) Nível fonético 2 (perceptivo) produzido pela escuta de sons dentro da palavra.
3) Nível fonêmico baseado em regras de pronúncia padrão.
4) Nível silábico que contempla a ressilabificação.
5) Nível de palavras fonéticas (grupos clíticos) com as marcações de acento.

6) Nível de entonação produzido por profissionais e baseado em análises auditivas e instrumentais.
7) Nível de *pitch* produzido automaticamente com posterior correção manual.
8) O nível das disfluências da fala e fenômenos não relacionados à fala é produzido manualmente para descrever os ruídos técnicos, hesitações, falsos começos, alongamentos, risos, etc.

Este esquema de anotação consome muito tempo, principalmente se for aplicado a um grande volume de dados. Dada uma tarefa específica de pesquisa, pode-se omitir alguns dos níveis apresentados. A pesquisa em entoação, por exemplo, não necessariamente exige a anotação completa: os segmentos são necessários apenas para calcular o alinhamento de pico, mas para isso precisa-se ter a informação apenas sobre as fronteiras de vogais acentuadas, e não de todos os segmentos de fala. O *corpus* de fala espontânea CoRuSS foi desenvolvido para a pesquisa em marcação automática de fronteiras prosódicas e, portanto, não inclui as fronteiras de sons de fala. No entanto, ainda contém a transcrição ortográfica e fonética – o que significa que poderemos adicionar as camadas que faltam mais tarde, quando os algoritmos para o alinhamento automático da fala forem capazes de fornecer uma precisão maior do que hoje em dia. Os nossos últimos teste mostraram um erro bastante elevado (em média, cerca de 20 ms), mas assim que atingirmos os números de erros significativamente mais baixos (pelo menos cerca de 6 ms), teremos a oportunidade de adicionar os níveis segmentais de alta qualidade aos *corpora* em que esses níveis estão em falta.

REFERÊNCIAS

Prilozhenie №3 k Byulletenyu Foneticheskogo fonda russkogo yazyka. Fond zvukovyh edinits russkoi rechi [Appendix # 3 to the Bulletin of the Russian Phonetic Fund]. Russian Phonetic Fund. St. Petersburg - Bochum, 1993.

BONDARKO, L. V.; SVETOZAROVA, N. D.; SKRELIN, P. A. *Foneticheskii fond russkogo yazyka kak issledovatel'skaya programma kafedry fonetiki Leningradskogo universiteta* [Russian Phonetic Fund as a Research Program of Department of Phonetics, Leningrad University]. Byulleten' Foneticheskogo fonda russkogo yazyka, St. Petersburg - Bochum, n.4, 1992.

BONDARKO, L. V.; VERBITSKAYA, L. A. (ed.) *Interferenciya zvukovyx sistem* [Cross-Language Influence of Sound Systems], Leningrad: Izdatel'stvo LGU, 1987.

BONDARKO, L. V.; VERBITSKAYA, L. A.; GORDINA, M. V.; KASEVICH, V. B. Stili proiznosheniya i tipy proizneseniya [Styles and Types of Pronunciation]. *Voprosy yazykoznaniya*, Moscow, n. 2. pp.64-70, 1974.

CHISTOVICH, L. A.; BONDARKO, L. V. Ob upravlenii artikulyatsionnymi organami v processe rechi [About Controlling Articulatory Organs in Speech Production]. *In: Issledovalia po strukturnoj tipologii* [Research in Structural Typology]. Moscow: Nauka, pp.169-182, 1963.

CHODROFF, E.; WILSON, C. Structure in Talker-Specific Phonetic Realization: Covariation of Stop Consonant VOT in American English. *Journal of Phonetics*, v. 61, pp.30-47, 2017.

EVDOKIMOVA, V.; EVGRAFOVA, K.; CHUKAEVA T. The Database of Normal and Pathological Singers' Voices: An Approach to Collecting Data. *In*: The 10th International Workshop Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA), 10., 2017, Florence. *Proceedings* [...]. Florence: Firenze University Press, 2017. pp.23-24.

EVDOKIMOVA, V.; KOCHAROV, D.; SKRELIN, P. Method for Constructing Formants for Studying Phonetic Characteristics of Vowels. *SPIIRAS Proceedings*, v. 19(2), pp.302-329, 2020.

EVDOKIMOVA, V.; SKRELIN, P.; CHUKAEVA, T. Automatic Phonetic Transcription for Russian: Speech Variability Modeling. *In:* International Conference on Speech and Computer (SPECOM), 19, 2017, Hatfield. *Proceedings* [...]. Springer International Publishing, 2017. pp.192-199.

EVDOKIMOVA, V.; ZAKHARCHENKO, E.; SKRELIN, P.; EVGRAFOVA, K.; CHUKAEVA, T.; SHVALEV, N. Akusticheskie xarakteristiki golosa v rechi i penii opernyx pevczov v norme i pri patologii [Acoustic Characteristics of Voice in Speech and Singing of Opera Singer's for Normal and Pathological Voice]. *In*: Interdisciplinary Seminar on Conversational Russian Speech Analysis, 8., 2019, Saint Petersburg. *Proceedings* […], Saint Petersburg: Polytechnika-print, 2019. pp.21-30.

EVGRAFOVA, K.; EVDOKIMOVA, V.; CHUKAEVA, T.; SKRELIN, P. Vocal Fatigue in Voice Professionals: Collecting Data and Acoustic Analysis. *In*: Tutorial and Research Workshop on Experimental Linguistics (EXLING 2016), 7., 2016, Saint-Petersburg. *Proceedings* [...], Saint-Petersburg: Saint Petersburg State University, 2016. pp.59-62.

GAROFOLO, J.; LAMEL, L.; FISHER, W.; FISCUS, J.; PALLETT, D.; DAHLGREN, N.; ZUE, V. *TIMIT Acoustic-Phonetic Continuous Speech Corpus*, 1993.

KACHKOVSKAIA, T.; CHUKAEVA, T.; EVDOKIMOVA, V.; KHOLIAVIN, P.; KRIAKINA, N.; KOCHAROV, D.; MAMUSHINA, A.; MENSHIKOVA, A.; ZIMINA, S. SibLing Corpus of Russian Dialogue Speech Designed for Research on Speech Entrainment. *In*: Conference on International Language Resources and Evaluation (LREC 2020), 12., Marseille. *Proceedings* [...], Marseille: ELRA, 2020. pp.6556-6561.

KACHKOVSKAIA, T.; KOCHAROV, D.; SKRELIN, P.; VOLSKAYA, N. CoRuSS—A New Prosodically Annotated Corpus of Russian Spontaneous Speech. *In*: Conference on International Language Resources and Evaluation (LREC 2016), 10., 2016, Portorož. *Proceedings* [...], Portorož: ELRA, 2016. pp.1949-1954.

KACHKOVSKAIA, T.; MAMUSHINA, A.; PORTNOVA, A. Typical and Rare Post-Nuclear Melodic Movements in Russian. *In*: Speech Prosody, 10., 2020, Tokyo. *Proceedings* [...], Tokyo: ISCA, 2020. pp.464-468.

KACHKOVSKAIA, T., MENSHIKOVA, A.; KOCHAROV, D.; KHOLIAVIN, P.; MAMUSHINA, A. Social and Situational Factors of Speaker Variability in Collaborative Dialogues. *In*: Speech Prosody, 11., 2022, Lisbon. *Proceedings* [...], Lisbon: ISCA, 2022. pp.455-459

KACHKOVSKAIA, T.; SKRELIN, P. Prosodic Phrasing in Russian Spontaneous and Read Speech: Evidence from Large Speech Corpora. *In*: Speech Prosody, 10., 2020, Tokyo. *Proceedings* [...], Tokyo: ISCA, 2020. pp.166-170.

KACHKOVSKAIA, T.; VOLSKAYA, N.; SKRELIN, P. Final Lengthening in Russian: A Corpus-Based Study. *In*: Interspeech 2013, 14., Lyon. *Proceedings* [...], Lyon: ISCA, 2013. pp.1438-1442.

KOCHAROV, D.; KACHKOVSKAIA, T.; SKRELIN, P. Prosodic Boundary Detection Using Syntactic and Acoustic Information. *Computer Speech and Language*, v. 53, pp.231-241, 2019a.

KOCHAROV, D.; KACHKOVSKAIA, T.;SKRELIN, P. Prosodic Factors Influencing Vowel Reduction in Russian. *In*: Interspeech 2019, 20., Graz. *Proceedings* [...], Graz: ISCA, 2019b. pp.1956-1960.

KOCHAROV, D.; VOLSKAYA, N.; SKRELIN, P. F0 Declination in Russian Revisited. *In*: International Congress of Phonetic Sciences (ICPHS), 18., 2015, Glasgow. *Proceedings* [...], Glasgow: International Phonetic Association, 2015.

KOCHETKOVA, U.; SKRELIN, P.; EVDOKIMOVA, V.; NOVOSELOVA, D. Perception of Irony in Speech. *In*: International Conference on Neurobiology of Speech and Language, 4., 2020, Saint Petersburg. *Proceedings* [...], Saint Petersburg: Skifia-Print, 2020. pp.72-73.

KOCHETKOVA, U.; SKRELIN, P.; EVDOKIMOVA, V.; NOVOSELOVA, D. The Speech Corpus for Studying Phonetic Properties of Irony. *In*: Language, Music and Gesture: Informational Crossroads, 2021, Saint Petersburg. *Proceedings* [...], Springer International Publishing, 2021. pp.203-214.

LIBERMAN, M. Corpus Phonetics. *Annual Review of Linguistics*, 5, pp.91-107, 2019.

MAKAROVA, V. A.; USENKOVA, E. V.; EVDOKIMOVA, V. V.; EVGRAFOVA, K. V. Yazyk saskachevanskix duxoborov: vvedenie v analiz [The Language of the Saskatchevan Doukhobors: Introduction and Analysis]. *Izvestiya vysshix uchebnyx zavedenij*. Seriya «Gumanitarnye nauki». Razdel lingvistika [New of Higher School. Humanities. Linguistics], Ivanovo, v. 2, n. 2, pp.146-152, 2011.

MENSHIKOVA, A.; KOCHAROV, D.; KACHKOVSKAIA, T. Phonetic Entrainment in Cooperative Dialogues: A Case of Russian. *In*: Interspeech 2020, 21., Shanghai. *Proceedings* [...], Shanghai: ISCA, 2020. pp.4148-4152.

O'CONNOR, J. D.; ARNOLD, G. F., *Intonation of Colloquial English*. Bristol, U.K.: Longman Group Ltd., 1973.

OSTENDORF, M.; PRICE P. J.; SHATTUCK-HUFNAGEL, S., *The Boston University Radio News Corpus*, Boston University Technical Report No. ECS-95-001, 1995.

PANAYOTOV, V.; CHEN, G.; POVEY, D.; KHUDANPUR, S. Librispeech: an ASR Corpus Based on Public Domain Audio Books. *In*: 2015 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 40., 2015, Brisbane. *Proceedings* [...], Brisbane: IEEE, pp.5206-5210.

PAPERNO, S.; LEED, R.L. Vocabulary Words in Elementary Russian Textbooks. *Slavic and Eats European languages journal*, v. 32, n. 2, 1988. pp.305-312

SKRELIN, P. Concatenative Russian Speech Synthesis: Sound Database Formation Principles. *In*: International Conference on Speech and Computer (SPECOM), 2., 1997. Cluj-Napoca. *Proceedings* [...], Cluj-Napoca: Editura Promedia Plus, 1997a.

SKRELIN, P. A. (ed.) Skazki Russkogo Severa [Tales of the North of Russia]. *In*: *Byulleten` foneticheskogo fonda russkogo yazyka. Prilozhenie 6* [The Bulletin of the Russian Phonetic Fund. Appendix 6]. Saint Petersburg - Bochum, 1997b.

SKRELIN, P. A. (ed.) Obryadovaya poeziya Russkogo Severa: plachi [Poetic Folklore of the North of Russia (Lamentations)]. *In*: *Byulleten` foneticheskogo fonda russkogo yazyka. Prilozhenie 6* [The Bulletin of the Russian Phonetic Fund. Appendix 6], Saint Petersburg - Bochum, 1998.

SKRELIN, P. A. *Segmentaciya i transkripciya* [Segmentation and Transcription], Saint Petersburg: Saint Petersburg State University, 1999.

SKRELIN, P. Russian Material and Methods. *In*: DE SILVA, V.; ULLAKONOJA, R. (ed.) *Phonetics of Russian and Finnish*, Frankfurt am Main: Peter Lang, 2009.

SKRELIN, P. A.; KOCHETKOVA, U. E.; EVDOKIMOVA, V. V.; NOVOSELOVA, D. D.; GERMAN, R. D. Prosodicheskie xarakteristiki ironicheskix vyskazyvanij v russkom i franczuzskom yazykax [Prosodic Features of Ironic Utterances in Russian and French]. *In*: Interdisciplinary Seminar on Conversational Russian Speech Analysis, 9., 2021, Saint Petersburg. *Proceedings* […], Saint Petersburg: Skifia-Print, 2021. pp.81-86.

SKRELIN, P.; VOLSKAYA, N.; KOCHAROV, D.; EVGRAFOVA, K.; GLOTOVA, O.; EVDOKIMOVA, V. A Fully Annotated Corpus of Russian Speech. *In*: Conference on International Language Resources and Evaluation (LREC 2010), 7., 2010, Valletta. *Proceedings* [...], Valletta: ELRA, 2010. pp.109-112.

SVETOZAROVA, N. Zhirmunsky's Collection of German Folk Songs in the Sound Archives of the Pushkinsky Dom. *In*: *Archives of the Languages of Russia*. Saint Petersburg - Groningen, 1996. pp.33-38.

TURK, A.; NAKAI, S.; SUGAHARA, M. Acoustic Segment Durations in Prosodic Research: A Practical Guide. Methods. *In*: *Empirical Prosody Research*, Berlin, Boston: De Gruyter, pp.1-28, 2012.

VOLSKAYA, N.; KACHKOVSKAIA, T. Prosodic Annotation in the New Corpus of Russian Spontaneous Speech CoRuSS. *In*: Speech Prosody, 8., 2016, Boston. *Proceedings* [...], Boston: ISCA, 2016. pp.917-921.

WANG, C.; RIVIERE, M.; LEE, A.; WU, A.; TALNIKAR, C.; HAZIZA, D.; WILLIAMSON, M.; PINO, J.; DUPOUX, E. VoxPopuli: A Large-Scale Multilingual Speech Corpus for Representation Learning, Semi-Supervised Learning and Interpretation. *In*: ACL 2021 (Volume 1: Long Papers), Bangkok, *Proceedings* […], Bangkok: ACL, 2021. pp.993-1003.

Traduzido por *Aleksandra S. Skorobogatova* as.skorobogatova@gmail.com

Recebido em 30/09/2021

Aprovado em 20/03/2023

Declaração de contribuição de autor

A contribuição de cada autor foi a seguinte:
Pavel Skrelin: Conceptualização dos princípios iniciais de definição do *corpora* de fala; evolução da metodologia para traçar o *corpora*; redação do rascunho original e revisão do artigo; revisão final e aprovação para publicação.
Tatiana Kachkovskaia: Evolução da metodologia de concepção de *corpora* de fala; redação do rascunho original, revisão e edição do artigo; revisão final e aprovação para publicação
Daniil Kocharov: Evolução da metodologia de concepção de *corpora* de fala; redação do rascunho original, revisão e edição do artigo; revisão final e aprovação para publicação
Vera Evdokimova: Evolução da metodologia de concepção de *corpora* de fala; Revisão final e aprovação da publicação.
Uliana Kochetkova: Evolução da metodologia de concepção de *corpora* de fala; Revisão final e aprovação da publicação.

Declaração de disponibilidade de conteúdo

Os conteúdos subjacentes ao texto da pesquisa estão contidos no manuscrito.

Pareceres

Tendo em vista o compromisso assumido por *Bakhtinina.* Revista de Estudos do Discurso com a Ciência Aberta, a revista publica somente os pareceres autorizados por todas as partes envolvidas.