Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference "Dialogue 2020"

Moscow, June 17–20, 2020

DISFLUENCIES IN RUSSIAN SPOKEN MONOLOGUES: A DISTRIBUTIONAL ANALYSIS[1]

Korotaev N. A. (n_korotaev@hotmail.com),
Podlesskaya V. I. (vi_podlesskaya@il-rggu.ru)
Russian State University for the Humanities, Moscow, Russia

Smirnova K. V. (kategold1@gmail.com),
Fedorova O. V. (olga.fedorova@msu.ru)
Lomonosov Moscow State University, Moscow, Russia

The paper addresses the overall distribution of speech disfluencies in Russian spoken monologic discourse: basing on corpus data, we investigate qualitatively and quantitatively how disfluencies of different types group (or do not group) with each other and how isolated disfluencies and their sequences are sandwiched with periods of fluent speech in the course of speech production. Self-repairs, filled and silent pauses, and instances of hesitation lengthening were annotated in a subcorpus of the "Russian Pears Chats and Stories" (RUPEX). A distribution-oriented typology of disfluencies was proposed that distinguishes between isolated disfluencies, disfluency clusters, and quasi-clusters. We claim that disfluency tokens tend to cluster, as isolated occurrences are significantly less frequent in our data than it could have been expected basing on the relative frequency of tokens. This finding contradicts previous studies that treated disfluency clusters as a more marginal phenomenon, and emphasizes the importance of a distributional, rather than merely structural, approach to annotating disfluencies. Furthermore, individual types of disfluency tokens demonstrate significantly different distributional patterns. Compared to other types, self-repairs occur more often in isolation, while words with hesitation lengthening appear predominantly in clusters, and filled pauses most often group with silent pauses to form quasi-clusters.

Key words: speech disfluencies, spoken discourse, corpus, disfluency clusters

DOI: 10.28995/2075-7182-2020-19-454-466

[1] This study is supported by Russian Foundation for Basic Research (complex grant #18-00-01598, and subgrant #18-00-01485).

ДИСТРУБИТИВНЫЙ АНАЛИЗ РЕЧЕВЫХ СБОЕВ В УСТНОМ РУССКОМ МОНОЛОГИЧЕСКОМ ДИСКУРСЕ

Коротаев Н. А. (n_korotaev@hotmail.com),
Подлесская В. И. (vi_podlesskaya@il-rggu.ru)
РГГУ, Москва, Россия

Смирнова К. В. (kategold1@gmail.com),
Федорова О. В. (olga.fedorova@msu.ru)
МГУ им. М. В. Ломоносова, Москва, Россия

В статье рассматриваются дистрибутивные свойства речевых сбоев в устном русском монологическом дискурсе. На материале корпуса «Рассказы и разговоры о грушах» мы проанализировали, как различные типы речевых сбоев группируются или не группируются друг с другом в процессе реального речепорождения. Этим наша работа отличается от большинства исследований речевых сбоев, в которых это явление чаще изучается не с синтагматических, а с парадигматических позиций. Мы рассмотрели четыре типа речевых сбоев: самоисправления, заполненные паузы, абсолютные паузы при хезитации и хезитационные удлинения. Все эти типы могут встречаться как изолированно, так и в составе кластеров или квази-кластеров (эти понятия определены в тексте статьи); причем в целом кластеризация оказывается более частотным явлением, чем изолированное употребление. Кроме того, согласно предварительному анализу, типы сбоев варьируют по степени стремления к кластеризации: самоисправления значимо чаще встречаются в изоляции, хезитационные удлинения — в составе кластеров, заполненные паузы — в составе квази-кластеров.

Ключевые слова: речевые затруднения, устная речь, корпус, кластеризация речевых сбоев

1. Introduction

Speech disfluencies have been studied under the auspices of various linguistic disciplines, including discourse analysis, conversational analysis, experimental psycholinguistics, speech recognition, speech pathology, etc. The state of the art in this area is well represented, inter alia, in the proceedings of biennial Workshops on Disfluency in Spontaneous Speech (e.g., in the latest [DiSS 2019]), in multiple dissertations (e.g., [Bailey 2004], [Schnadt 2009], [Miller 2010], [Finlayson 2014]) and overviews ([Lickley 2015] is among the best), see also The Filled Pause Research Center—a very useful open online resource at http://filledpause.com.

Beginning with seminal works by [Willem J. M. Levelt 1983], [Elizabeth Shriberg 1994] and [Robert Eklund 2004], studies on disfluencies addressed primarily the typology of disfluencies as well as forms and functions of their individual types in individual

languages (cf. [Watanabe 2006], [2013] on Japanese, [Gósy 2019] on Hungarian, [Silber-Varod et al. 2019] on Hebrew, [Plug 2015] on Dutch, etc.) The most articulated research questions cover (a) the localization of disfluencies against hierarchical and linear structure of discourse, i.e. whether disfluencies appear at the boundaries of words, syntactic units, intonation units, turns etc., or inside the respective unit; (b) the influence of discourse genres (narratives, argumentative discourse, descriptions, etc.) and types of discourse (monologue vs. dialogue) on forms and frequencies of disfluencies; (c) segmental and suprasegmental characteristics of disfluent speech in comparison with fluent standards. In addition to these, an acute discussion in the literature on disfluencies, initiated primarily in a salient paper by [Herbert H. Clark and Jean E. Fox Tree 2002], concerns the functions of particular disfluencies (mainly, filled pauses) from the perspective of listeners: the central research question here is whether listeners comprehend disfluencies as intentional signals produced by speakers to inform about speech difficulties or as involuntary symptoms of those (cf. [Finlayson 2014] among recent contributions to this discussion).

However, much less is known and much less discussed is how disfluencies are distributed throughout big coherent portions of discourse; namely, how fluent and disfluent portions of speech alternate each other, which types of difluences cluster more easily, and which appear as isolated tokens surrounded by fluent speech? Literature on these distributional issues is much more scarce, some few exceptions to be mentioned include e.g. [Crible et al. 2017], [Betz et al. 2015], [Betz and Kosmala 2019], [Bóna 2019]. Our paper is in line with this distributional approach, which we advocate and check against Russian data. Since Russian is considerably underrepresented in disfluency studies, we aim at partially bridging also this gap.

Our main goal is to explore the overall distribution of speech disfluencies in Russian spoken monologic discourse. Basing on corpus data, we investigate qualitatively and quantitatively how disfluencies of different types group (or do not group) with each other and how isolated disfluencies and their sequences are sandwiched with periods of fluent speech in the course of speech production.

The following example may serve as a preliminary illustration of some (though not all!) phenomena we deal with in this study, including self-repairs as well as filled and silent pauses. These disfluencies can appear as singletons surrounded by fluent material or can group together.

(1) pears04: N-vE035—N-vE040[2]

N-vE035	/Podnimaetsja climbs.up	\nazad. back	
pN-016	(0.30)		
N-vE036	Prodolžaet continues	sobirat' to.pick	\↑gruši[u], == pears

[2] See **Section 2** for an overview of our data and transcripts details, and **Section 3** for additional glossing conventions.

N-vE037	(*\Da̲,								
	yes								
N-vE038	v	/kakoj-to	moment	proxo=	||	(ə 0.25)	\bleet	–/koza̲,	
	in	some	moment	passes.b=	BR	FP	bleats	goat	
N-vE039	my	slyšim	/↑ble̲janie,						
	we	hear	bleating						
pN-017	(0.23)								
	SP								
N-vE040	(əɯ 0.58)		idët	/čelove̲k,					
	FP		walks	man					

'He climbs back up [the tree]. He resumes picking the pears... Oh yes, at some point a goat is pass-… bleating; we can hear the bleating and [we see] a man walking'

In (1), the first discourse unit, i.e. the first line in the transcript, is produced fluently, then comes a silent pause. At the end of the next line, N-vE036, the speaker abandons the current discourse unit and makes a switch in the narration line, as she suddenly decides to add a missing detail about a man with a goat who passes by (this type of repair, which affects the overall structure of a relatively big discourse fragment, is notated by a double equal sign). In line N-vE038, two disfluencies are grouped together: first, the word *proxo=* (initially intended as *proxodit* 'passes by') is truncated and further on substituted by the word *bleet* 'bleats' (this type of repair—caused by word searching difficulties inside the ongoing discourse unit—is notated by double vertical strokes); and then, immediately after that, comes a filled pause *(ə)*. Later on, another filled pause, *(əɯ)*, is found at the beginning of line N-vE040; notably, it is immediately preceded by a silent pause. We will refer to example (1) later, in **Section 3**.

The rest of the paper is organized as follows. In **Section 2**, we present our data. In **Section 3**, we discuss the principles for annotating disfluencies, propose a distributive typology thereof, and analyze some typical examples. **Section 4** contains quantitative results and their statistical evaluation. In **Section 5**, we provide a conclusion.

2. Data

Our data come from the "Russian Pear Chats and Stories" corpus (RUPEX; see [Kibrik, Fedorova 2018]). RUPEX consists of annotated communication sessions that share the same general design. In each session, three participants with fixed roles (Narrator, Commentator, and Reteller) alternate in relating and discussing the contents of the "Pear Film" [Chafe (ed.) 1980]. RUPEX is a multichannel corpus; it registers not only speech, but also kinetic actions such as manual gesticulation, head movements, and gaze direction. The analysis in this study, however, is confined to vocal modality. We used the ***vocal transcripts*** available on RUPEX website (https://multidiscourse.ru/corpus/?en=1). In transcripts, each line corresponds to an elementary discourse unit (EDU; see [Kibrik, Podlesskaya (eds.) 2009]) or a pause between EDUs; pause durations are indicated in parentheses; accents placement and pitch

directions are iconically indicated with slashes and arrows; punctuation marks at the end of EDUs code illocutionary and phase values; see [Korotaev et al. 2020] for more detail. Additional conventions are discussed in this paper, when needed.

Three recorded sessions were analyzed for this study. In examples, we provide the session ID (04, 22, or 23) and IDs of the EDUs involved (in these IDs, N stands for Narrator, C for Commentator, and R for Reteller). Since we were primarily interested in how fluent and disfluent stretches are distributed across relatively long speech intervals, we concentrated on monologic fragments and didn't consider dialogical parts[3]. Overall, we annotated 32 mins of audio that contained 4,780 words.

3. Annotating disfluencies

When annotating disfluencies, we implemented a ***distribution-oriented typology*** that draws a primary distinction between isolated disfluencies and disfluency clusters. ***Isolated*** are single disfluency tokens surrounded by fluent stretches; ***clusters*** are sequences of disfluency tokens that come together. Before discussing this distinction in more detail (see **Section 3.2**), we will first present the types of tokens that were taken into account.

3.1. Basic disfluency tokens

Basic disfluency tokens are what most traditional classifications are about; they are elementary signals (or symptoms[4]) of speech difficulties. Since our analysis was primarily aimed at revealing the distributional patterns of disfluencies, we used a simplified typology of basic tokens that includes the following options.

3.1.1. Self-repairs

In self-repairs, speakers interrupt their speech, reject an already pronounced fragment, and then restart or correct it. Following the classical model introduced in [Schriberg 1994], we analyze self-repairs as having three elements: reparandum, interruption point (often signaled by words truncation), and repair (or, reparans). In example (1) provided above, there are two instances of self-repair, one at a macro-level, and the other one—which is combined with a filled pause—inside an EDU. Another instance of a self-repair inside an EDU is found in (2). Here, a so-called false-start takes place. As the speaker starts uttering the word *nežno* 'gently', she decides that it should not precede but rather follow the verb *delaet* 'does'; so she immediately interrupts her production [Levelt, 1993: 478], provides the verb, and then restarts the suspended verbalization. In examples provided in this paper, we gloss interruption points as BR (from "break").

---

[3] In our prior work, we found that monologues and dialogues differ in what speech disfluencies occur more frequently. Compared to monologues, speakers in dialogues tend to produce significantly less filled pauses [Korotaev 2019] and more repairs [Podlesskya et al. 2019]. This is another reason why we decided to stick to the monologic data in our current research.

[4] See **Section 1** for a brief overview of this delimitation in the literature. As for now, we are not ready to adhere to a particular decision as to whether speech disfluencies are rather signals of difficulties in planning, or mere (non-voluntary?) symptoms thereof. In this paper, we try to abstract ourselves from this distinction and use somewhat neutral terms.

(2) pears04: N-vE034

N-vE034 tak vsë očen' ne= || delaet /nežno i \ljubovnoʷ.
this.way everything very tend= BR does tenderly and lovingly

'And he is tend-… doing all this in a very tender and loving way'

Excerpts (1) and (2) illustrate different types of self-repairs. Numerous classifications have been brought to account on these and still other differences, including the scope and the localization of the repair, the type of the repaired feature (lexical, grammatical, constructional), etc.; see, inter alia, [Schegloff 2013]; [Podlesskaya 2015]. In this study, for the sake of simplicity, we do not distinguish between different types of self-repairs and treat them all uniformly. Self-repair is a complex phenomenon that involves several parts; in the quantitative analysis below, cf. **Section 4**, counted are frequencies of interruption points.

3.1.2. Filled pauses (FPs)

There are two instances of filled pauses in example (1) above, one combined with a self-repair, the other following a silent pause. An example of an FP surrounded by fluent speech is given in (3), where it occurs between the verb and the adjunct.

(3) pears22: R-vE163

R-vE163 i /gruši razletelis' (ə 0.49) \vokrug.
and pears got.scattered FP around

'And the pears got scattered all around'

Filled pauses may have slightly different functions depending on their phonetic realizations [Clark, Fox Tree 2002], but their general role of hesitation markers seems undisputable. In all the examples above, an FP is used when the speaker hesitates on how to continue her production and seeks to gain some additional time[5].

3.1.3. Lengthening

Functionally and formally close to FPs are some cases of lengthening, or phoneme prolongation. In (4), the speaker prolongates the final vowel of the complex preposition *v storonu* 'towards' before providing the noun *dereva* 'tree'. A plausible interpretation of this lengthening is that the speaker first seeks for a more concrete nomination (like, for instance, 'a pear tree') but then rejects this idea and uses a generic term. In transcripts, lengthening is conveyed with a hyphenated notation.

(4) pears22: R-vE207

R-vE207 poka značit ᵃmal'čiki idut v storonu-u \ↅdereva,
while so boys go towards tree

'So, while the boys are heading towards the tree…'

[5] In RUPEX, several types of FPs are distinguished—(*ə*) and (*ɐ*) for *uh*- and *ah*-like pauses, (*ɯ*) for *um*-like pauses, (ˀ) for glottal stop fillers, as well as various mixed cases (see (*əm*) in (1)). In our analysis, we neglect these distinctions.

As has been repeatedly pointed out in literature, in cases like (4) lengthening is similar to filled pauses as it also signals hesitation by means of extended vocalization (see, inter alia, [Eklund 2004: 241–252]). However, there are numerous uses of this formal device beyond hesitation. Lengthening may express emphasis [Braver et al. 2014], intensification and spatio-temporal relations [Kodzasov 2009: 112–125, 137–146], appear in phonetically motivated contexts [Wightman et al. 1992], take part in complex discourse strategies such as addressing someone or remembering something [Yanko 2008: 98–117], and so on. In this study, we annotated prolongations as disfluency tokens only when we were quite sure of their hesitation function in the given context. As for now, we are not aware of ways to make this procedure more formal.

3.1.4. Silent pauses (SPs)

Similar problems arise when annotating silent, or absolute, pauses. Sometimes, when an SP appears in an "unexpected" position and/or is "unexpectedly" long, it can also be interpreted as a hesitation phenomenon. For instance, in (5) the speaker suspends vocalization after providing the adverb *tuda* 'there', keeps silent for more than 600 ms, and then completes the clause by producing the verb *skladyvaet* 'puts'. This can hardly be interpreted otherwise than an instance of hesitation, as the speaker is apparently trying to find a good way to express the idea of putting an already indicated referent into an indicated place.

(5) pears23: R-vE156

R-vE156	on	èti	/gruši	tuda	(0.61)	\skl̲adyvaet.
	he	these	pears	there	SP	puts

'He puts these pears there'

Still, just like lengthening, SPs don't only occur as hesitation devices. Most importantly, silent pauses, as well as pauses filled with loud inhalations[6], tend to coincide with boundaries between elementary discourse units and other discourse constituents (see [Selkirk 1984]; [Chafe 1994]; [Kibrik, Podlesskaya (eds.) 2009]; etc.). Perception-based criteria are usually used to tell these two functions apart (see, e.g., [Lickley 2015: 458]; [Trouvain et al. 2016]); and we followed the same principle. Here, it suffices to say that the SP in (5) was treated as a disfluency token, while a shorter pause placed between lines N-vE035 and N-vE036 in (1) was not.

3.1.5. Other

There are still other markers of disfluency: placeholders (*ètot* 'that one', *kak ego* 'whatchacallit'), lexical search items (*nu* 'well'), editing terms (*to est'* 'that is', *net* 'no'), etc. They were not annotated for this study.

3.2. Isolated vs. clustered disfluencies

As noted above, the distribution-oriented typology we propose in this study is based on the distinction between isolated disfluencies and disfluency clusters. Disfluencies

[6] SPs per se and those filled with loud inhalations are annotated separately in the corpus but are not differentiated for counts in this study.

presented in examples (2)–(5) were annotated as isolated since they are surrounded by fragments of fluent speech. Contrary to these cases, clusters are sequences of disfluency tokens that come adjacent to one another. For example, in line N-vE038 of excerpt (1) above, the interruption point of a self-repair (truncated verb *proxo=* 'passes b=') is immediately followed by an FP (*ə 0.25*). A more complex cluster is present in (6). The line starts with an FP, then the final vowel of the pronoun *oni* 'they' is prolongated and gradually evolves into another FP, after that the speaker truncates her first attempt to provide the verb *razminajutsja* 'pass by each other' and the interruption point is once again followed by an FP. Overall, five adjacent disfluency tokens precede a stretch of fluent speech.

(6) pears04: N-vE097

N-vE097	(ɐ 0.30)	Oni-iᵊ	(ɐ 0.36)	r=	||	(ə 0.23)	razminajutsja		na	–↑dorogeᵚ,
	FP	they	FP	p-	BR	FP	pass.by.each.other		on	road

'They pass by each other on the road'

We consider the opposition of isolated vs. clustered disfluencies significant, as the latter type presumably indicates more severe difficulties than the former (see, e.g., [Robb et al. 2009]; [Bóna 2019] for clinical accounts on clustering). However, there is no clear-cut between these two types. We find it useful to point out at least two intermediate cases. First, sometimes disfluency tokens don't follow one another immediately, but are still apparently connected. In (7), the speaker enters the "disfluency mode" as she prolongates the final consonant of *potom* 'then' in line N-vE083. After that, she successfully provides the matrix verb *ponimaet* 'understands' but resumes hesitating on what to say next and uses a combination of lengthening and FP right in the beginning of the complement clause.

(7) pears04: N-vE083

N-vE082	snačala	mal'čik	xočet	vzjat'	/odnu	↑grušu,			
	at.first	boy	wants	to.take	one	pear			
N-vN023	(ɥ 0.49)								
N-vE083	/potom-m	ponimaet	čto-o	(ʔ 0.45)	ničto	emu	ne	/grozit,	
	then	understands	that	FP	nothing	to.him	not	threatens	

'At first, the boy wants to take only one pear, but then he understands that he's in no danger…'

Cases like (7) were also annotated as clusters. To discriminate between clusters and isolated disfluencies we used a one-word threshold: tokens separated by no more than one fluent notional word were considered elements of clusters[7].

Second, quite frequent (see **Section 4**) are sequences of single disfluency tokens immediately followed and/or preceded by silent pauses. Such sequences were annotated as ***quasi-clusters***; see (8), where an FP is preceded by an SP. Occurring between the topical noun phrase and the rhematic verb phrase, the SP doesn't indicate

[7] A looser, two-word, threshold is used in [Gósy 2012].

disfluency in itself, but followed by an FP, it may be reinterpreted as a hesitation device. (A similar quasi-cluster is present in lines pN-017—N-vE040 of excerpt (1).)

(8) pears04: R-vE497

R-vE497 V èto /vremja /–fermer (0.55) (ə 0.31) spuskaetsja s /↓→lestnicy„„
in this time farmer SP FP goes.down from ladder

'At this point, the farmer gets down the ladder'

4. Results and discussion

440 disfluencies were annotated in our data. Table 1 shows the distribution of these disfluencies across isolated, clustered, and quasi-clustered types.

Table 1. Isolated disfluencies, disfluency clusters, and quasi-clusters in annotated data

	Isolated	Quasi-clusters	Clusters	Overall
N (%)	126 (28.6%)	174 (39.6%)	140 (31.8%)	440 (100%)

Numbers in Table 1 suggest that, contrary to some previous studies (see, e.g., [Betz, Cosmala 2019]), clustering of disfluencies is not a marginal phenomenon, but rather a standard tendency. Even if we don't consider quasi-clusters (the most frequent, and somewhat intermediate, case), proper clusters still occur more frequently than isolated disfluencies in our data.

To provide a preliminary statistical evaluation of these findings, we analyzed the distribution of disfluency tokens across isolated and clustered disfluencies. Our assumption was that tokens don't appear independently of one another, i.e. that clustering is not a matter of coincidence. For the sake of simplicity, we eliminated all silent pauses from the analysis, so tokens that appear in quasi-clusters were counted as isolated. The results are as follows. Out of 672 disfluency tokens (self-repairs, FPs, words with lengthening), 391 are used in clusters, and 281 are isolated. We compared the observed distribution with the distribution expected under the null hypothesis, as shown in Table 2. The expected distribution was calculated based on the relative frequency of disfluency tokens in the analyzed data (672 / 4,780 words; i.e. 0.14). If disfluency tokens appeared independently of one another, the probability for a word to have at least one disfluency token at a 2-word distance (see discussion in Section 3.2) would equal 0.45 ($1 - 0.86\hat{}4$, as the distance should be laid out in both directions). The difference between the observed and the expected distribution is significant (chi-square test; $p < .00001$); the null hypothesis was rejected.

Furthermore, we observed significant differences in how specific types of disfluency tokens comply with the overall tendency of clustering. Fig. 1 presents the normalized distributions of filled pauses, words with lengthening, and self-repairs across isolated, quasi-clustered, and clustered disfluencies.

Table 2. Observed vs. expected distributions of disfluency tokens across isolated and clustered types

Tokens occurring:	Observed	Expected
as isolated disfluencies or in quasi-clusters	281	370
in clusters	391	302

Figure 1. Appearance of disfluency tokens in isolated, quasi-clustered, and clustered disfluencies

The differences shown in Fig. 1 are statistically significant (chi-square test; $p < .00001$). While all three types are predominantly used in clusters, self-repairs seem to be the most independent type of disfluencies, as 29.9% of them (50 out of 167; remember that we count interruption points here) occur in isolation. Filled pauses, on the contrary, are isolated in just 8.3% (27 out of 325), but often (38.8%) appear in quasi-clusters, i.e. come exclusively grouped with silent pauses and no other disfluency tokens. As for lengthening, tokens of this type most frequently appear in clusters—71.8% (128 out of 180), compared to 52.9% and 54.5% of filled pauses and self-repairs, respectively.

5. Conclusions

We analyzed the speech disfluencies (self-repairs, filled pauses, hesitation lengthening and silent pauses) found in the monologic parts of three communicative sessions included into the "Russian Pear and Chats Stories" corpus. Our main concern was to find out whether disfluencies tend to group together, or they most frequently appear in isolation of one another. We annotated our data using a distribution-oriented typology of disfluencies that distinguishes between isolated cases, clusters, and quasi-clusters. Clusters turned out to be quite frequent, as they accounted for nearly 32% of all

disfluencies. This allowed us to assume that clustering, often considered as manifesting especially severe speech difficulties, is in fact a standard feature of natural spoken discourse. We provided a preliminary statistical evaluation of this claim by further investigating into how often basic disfluency tokens (filled pauses, lengthening, and self-repairs) group or do not group together. We found that over 58% of 672 tokens appeared in clusters, which is way more often than it could have been expected basing on the relative frequency of tokens in our data. Also, specific types of disfluency tokens behave somewhat differently, as self-repairs occur more often in isolation, while words with hesitation lengthening appear predominantly in clusters, and filled pauses most often group with silent pauses to form quasi-clusters.

We foresee several expansions of our study. First, more accurate results on clustering could be obtained when all instances of disfluency tokens are annotated separately. As we concentrated on a holistic analysis of disfluencies, this has not been properly done yet for our data. Second, a typology of clusters could be introduced basing on number, types, and respective order of disfluency tokens. For example, we found that clusters containing one or several filled pause(s) and one or several instance(s) of lengthening are the most frequent in our data (50 out of 140 clusters, or 35.7%); a finer-grained annotation could yield more specific results. Finally, we understand that our methodology could be applied for annotating other related phenomena like speech disfluencies in people who stutter or clustering in manual gesticulation. Also, it has already been used for selecting stimuli in an event-related fMRI experiment [Smirnova et al. 2020].

Acknowledgments

We express our gratitude to Svetlana Toldova for her help with a statistical assessment of our data. We also thank anonymous reviewers for their insightful observations. All remaining mistakes are, of course, our responsibility.

References

1. *Bailey K. G. D.* (2004), Disfluent speech and the visual world: An application of the visual world paradigm to the study of spoken language comprehension, PhD dissertation, Michigan State University, USA.
2. *Betz S., Wagner P., Schlangen D.* (2015), Micro-Structure of Disfluencies: Basics for Conversational Speech Synthesis, Proceedings of Interspeech 2015, Dresden, Germany, pp. 2222–2226.
3. *Betz S., Kosmala L.* (2019), Fill the silence! Basics for modeling hesitation, Proceedings of DiSS 2019, The 9th Workshop on Disfluency in Spontaneous Speech 2019, Eötvös Loránd University, Budapest, Hungary, pp. 11–14.
4. *Bóna J.* (2019), Clustering of disfluencies in typical, fast and cluttered speech, Clinical Linguistics & Phonetics, Vol. 33, Iss. 5, pp. 393–405.
5. *Braver A., Dresher N., Kawahara S.* (2014), The Phonetics of Emphatic Vowel Lengthening in English, Proceedings of the 2014 Annual Meeting on Phonology, available at https://journals.linguisticsociety.org/proceedings/index.php/amphonology/article/view/3754/3473.

6. *Chafe W.* (Ed.) (1980), The pear stories: Cognitive, cultural, and linguistic aspects of narrative production, Norwood: Ablex.
7. *Chafe W.* (1994), Discourse, consciousness, and time, University of Chicago Press, Chicago.
8. *Clark, H. H., Fox Tree J. E.* (2002), Using uh and um in spontaneous speaking, Cognition, Vol. 84, pp. 73–111.
9. *Crible L., Degand L., Gilquin G.* (2017), The clustering of discourse markers and filled pauses: A corpus-based French-English study of (dis)fluency, Languages in Contrast, Vol. 17, Iss. 1, pp. 69–95.
10. *Corley M., Stewart O. W.* (2008), Hesitation disfluencies in spontaneous speech: The meaning of um, Language and Linguistics Compass, Vol. 2, Iss. 4, pp. 589–602.
11. *Eklund R.* (2004), Disfluency in Swedish human-human and human-machine travel booking dialogues, PhD dissertation, University of Linkoping, Sweden.
12. *Finlayson I. R.* (2014), Testing the roles of disfluency and rate of speech in the coordination of conversation, PhD dissertation, Queen Margaret University, Edinburgh, UK.
13. *Gósy M.* (2012), BEA—A multifunctional Hungarian spoken language database, The Phonetician, Vol. 105, Iss. 10, pp. 50–61.
14. *Gósy M.* (2019), Halt command in word retrieval, Proceedings of DiSS 2019, The 9th Workshop on Disfluency in Spontaneous Speech, Eötvös Loránd University, Budapest, Hungary, pp. 3–6.
15. *Kibrik A. A., Fedorova O. V.* (2018), Language production and comprehension in face-to-face multichannel communication, Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog 2018" [Komp'juternaya Lingvistika i Intellektual'nye Texnologii: Trudy Mežunarodnoj Konferencii "Dialog 2018"], Moscow, RGGU, pp. 305–316.
16. *Kibrik A. A., Podlesskaya V. I.* (eds.) (2009), Night Dream Stories: A corpus study of spoken Russian discourse [Rasskazy o snovidenijax: Korpusnoe issledovanie ustnogo russkogo diskursa], LRC, Moscow.
17. *Kodzasov S. V.* (2009), Studies in Russian Prosody [Issledovanija v oblasti russkoj prosodii], LRC, Moscow.
18. *Korotaev N. A.* (2019), Hesitation pauses in narratives and conversations: A quantitative comparison [Pausy xesitacii v rasskaze i v razgovore], Proceedings of the Corpus Linguistics Conference 2019, [Trudy meždunarodnoj konferencii "Korpusnaja lingvistika-2019"], Saint Petersburg University, Saint Petersburg, pp. 48–54.
19. *Korotaev N. A., Kibrik A. A., Podlesskaya V. I.* (2020), Annotating the vocal modality, O. V. Fedorova, A. A. Kibrik (eds.), The MCD handbook: A practical guide to annotating multichannel discourse, in print.
20. *Levelt W. J. M.* (1983), Monitoring and self-repair in speech, Cognition, Vol. 14, Iss. 1, pp. 41–104, 1983.
21. *Levelt W. J. M.* (1993), Speaking: From Intention to Articulation, The MIT Press.
22. *Lickley R. J.* (2015), Fluency and Disfluency, Melissa A. Redford (ed.) The Handbook of Speech Production, John Wiley & Sons, pp. 445–474.
23. *Miller T. A.* (2010), Generative models of disfluency, PhD dissertation, University of Minnesota.
24. *Plug L.* (2015), Prosodic Marking and Predictability in Lexical Self-Repair, Proceedings of the 18th International Congress of Phonetic Sciences 2015, Glasgow, UK, available at https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0032.pdf.

25. *Podlesskaya V. I.* (2015), A corpus-based study of self-repairs in Russian spoken monologues, Russian Linguistics, Vol. 39, Iss.1, pp. 63–79.
26. *Podlesskaya V. I., Korotaev N. A., Mazurina S. I.* (2019), A corpus study of self-repairs in Russian monologues and dialogues [Samoispravlenija govorjaščego v russkom monologičeskom i dialogičeskom diskurse: opyt korpusnogo issledovanija], Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2019" [Komp'juternaja Lingvistika i Intellektual'nye Texnologii: Trudy Meždunarodnoj Konferencii "Dialog 2019"], RGGU, Moscow, pp. 508–522.
27. *Robb M. P., Sargent A., O'Beirne G. A.* (2009), Characteristics of disfluency clusters in adults who stutter, Logoped Phoniatr Vocol., Vol. 34, Iss. 1, pp. 36–42.
28. *Schegloff E. A.* (2013), Ten operations in self-initiated, same-turn repair, Makoto Hayashi, Geoffrey Raymond, Jack Sidnell (Eds.), Conversational repair and human understanding, Cambridge: Cambridge University Press, pp. 41–70.
29. *Schnadt M.* (2009), Lexical influences on disfluency production, PhD dissertation, University of Edinburgh, UK.
30. *Selkirk E. O.* (1984), Phonology and Syntax: The Relation Between Sound and Structure, Cambridge, MA: The MIT Press.
31. *Shriberg E.* (1994), Preliminaries to a theory of speech disfluencies, PhD dissertation, University of California, Berkeley, USA.
32. *Silber-Varod V., Gósy M., Eklund R.* (2019), Segment prolongation in Hebrew, Proceedings of DiSS 2019, The 9th Workshop on Disfluency in Spontaneous Speech 2019, Budapest, Hungary, pp. 47–50.
33. *Smirnova K. V., Korotaev N. A., Panikratova Y. R., Lebedeva I. S., Pechenkova E. V., Fedorova O. V.* (2020), Using the RUPEX Multichannel Corpus in a Pilot fMRI Study on Speech Disfluencies, Twelfth International Conference on Language Resources and Evaluation (LREC), pp. 195–203, available at http://www.lrec-conf.org/proceedings/lrec2020/pdf/2020.lrec-1.25.pdf.
34. *Trouvain J., Fauth C., Möbius B.* (2016), Breath and Non-breath Pauses in Fluent and Disfluent Phases of German and French L1 and L2 Read Speech, Proceedings of Speech Prosody (SP8) 2016, Boston, USA, pp. 31–35.
35. *Watanabe M., Hirose K., Den Y., Miwa S., Minematsu N.* (2006), Factors influencing ratios of filled pauses at clause boundaries in Japanese, Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics 2006, Athens, Greece, pp. 253–256.
36. *Watanabe M.* (2013), Phrasal complexity and the occurrence of filled pauses in presentation speeches in Japanese, Proceedings of DiSS 2013, the 6th Workshop on Disfluency in Spontaneous Speech and TMH-QPSR 2013, KTH Royal Institute of Technology, Stockholm, Sweden, Vol. 54, Iss. 1, pp. 71–72.
37. *Wightman C. W., Shattuck-Hufnagel S., Ostendorf M., Price P. J.* (1992), Segmental durations in the vicinity of prosodic phrase boundaries, The Journal of the Acoustical Society of America, Vol. 91, Iss. 3, pp. 1707–1717.
38. *Yanko T. E.* (2008), Intonation strategies of Russian speech from a contrastive perspective [Intonacionnye strategii russkoj reči v sopostavitel'nom aspekte], LRC, Moscow.