

REPEATS IN ADVANCED SPOKEN ENGLISH OF LEARNERS WITH CZECH AS L1

TOMÁŠ GRÁF

ABSTRACT

The article reports on the findings of an empirical study of the use of repeats – as one of the markers of disfluency – in advanced learner English and contributes to the study of L2 fluency. An analysis of 13 hours of recordings of interviews with 50 advanced learners of English with Czech as L1 revealed 1,905 instances of repeats which mainly (78%) consisted of one-word repeats occurring at the beginning of clauses and constituents. Two-word repeats were less frequent (19%) but appeared in the same positions within the utterances. Longer repeats are much rarer (<2.5%). A comparison with available analyses show that Czech advanced learners of English use repeats in a similar way as advanced learners of English with a different L1 and also as native speakers. If repeats are accepted as fluencemes, i.e. components contributing to fluency, it would appear clear that many advanced learners either successfully adopt this native-like strategy either as a result of exposure to native speech or as transfer from their L1s. Whilst a question remains whether such fluency enhancing strategies ought to become part of L2 instruction, it is argued that spoken learner corpora also ought to include samples of the learners' L1 production.

Key words: fluency, disfluency, repeats, repetitions, L2 fluency, fluencemes

1. Introduction

The continuous flow of spontaneous speech production is frequently patterned with performance phenomena which include especially lexical and non-lexical fillers, pauses, draws, truncations, false starts, self-corrections, editing expressions and repeats. These phenomena are understood to relieve the pressure of online planning as their production helps the speaker to acquire time for planning in order to align mental planning with the physical aspects of speech production, or as is frequently the case for L2 speakers for choosing the appropriate form for the message which is being relayed. Unless these elements are too audible (e.g. loud filled pauses) or unusually frequent, they often go unnoticed and do not disturb the listener, who might, in fact, be informed by their pres-

ence that the speaker is aiming to carry on talking or that he is in the process of finding the desired content or form. However, as these features are semantically superfluous to the overall utterance they are generally labelled as disfluencies or dysfluencies and thus carry a rather negative connotation, being seen as elements which disrupt speech fluency.

In a seminal study of hesitation phenomena, Maclay and Osgood (1959) refer to disfluencies as “hesitation errors” and to those who produce fewer of them as “better speakers” (p. 35). As research develops, disfluencies come to be seen not only less negatively but also as essential and natural components of speech production. Fox Tree and Clark (1997) assume that they present a strategy whose function is to solve processing difficulties (see also Clark, 2002). Such a view is justifiable if we consider how ubiquitous disfluencies are (Biber et al., 1999; Kjellmer, 2008). Clark and Wasow (1998) link disfluencies with planning problems and explore them as evidence of planning. This view is developed by Segalowitz (2010) who analyses Levelt’s speech production model¹ and identifies within it seven “vulnerability points for fluency” (p. 9). These are defined as “critical points where underlying processing difficulties could be associated with L2 speech dysfluencies”. Segalowitz’s approach presents a deep-structure model for disfluencies in that he does not provide analyses of concrete instances of disfluencies but instead focuses on identifying where in the model problems may occur. As it is based on Levelt, it is not language-specific (we do not know how disfluencies are realised in different languages) and does not offer a surface-structure view which would investigate concrete realisations of these problems in terms of their qualitative, temporal, or locational characteristics. The nature of the problems is primarily in encoding on grammatical, lexical or phonological level. Segalowitz thus treats disfluencies as predominantly hesitational in nature, but he also acknowledges the existence of vulnerability points in the conceptual preparation phase and in the self-perception processes (see also Li and Tilsen’s (2015) discussion of whether disfluencies stem from planning or monitoring).

Skehan (2003) offers a more surface-structure view in his tri-partite model of fluency which sees speech fluency as a sum of speed fluency, breakdown fluency and repair fluency. In this model, some disfluencies are of hesitational nature, while others, such as repeats, aim to repair what has broken down in order to restore the impression of continuous speech. They are used as a communicative strategy, and as is suggested by Rühlemann (2006) should not be called dysfluencies – as dys- implies abnormality – but rather disfluencies, using a weaker reversative prefix. Götz (2013) goes even further in introducing the concept of a fluenceme, which is any component of speech which contributes to either productive or perceptive fluency. In her model of fluency, those features traditionally labelled as dysfluent or disfluent are categorized alongside such phenomena as speech rate or n-grams and rather than hesitational are seen as strategic. Such a view, however, fails to acknowledge that not all disfluent behaviour is necessarily strategic.

Whilst disfluencies have a role in helping the speaker to formulate his message, they also have an effect on the recipient and on the process of comprehension. MacGregor et al. (2009) show that in this respect not all disfluencies are the same: filled pauses are processed with greater ease while repeats are more disruptive as structural and semantic

¹ Levelt, W. J. M. (1999). Language Production: A Blueprint of the Speaker. In: Brown, C. & Hagoort, P. (Eds.), *Neurocognition of Language*, 83–122. Oxford: Oxford University Press.

interpretation must be restarted. This may be especially true of non-native-speaker L2 processing as was shown by Voss (1979) who found that hesitation phenomena were sources of perceptual errors and problems for non-native speakers.

Despite their ubiquitous nature, the production of disfluencies may vary from speaker to speaker. This may give rise to different patterns of disfluent behaviour with differences in the type of disfluency used, its frequency or different combinations of disfluent elements. Disfluent behaviour is thus, to a certain extent, seen as speaker-specific. This was observed already by Maclay and Osgood (1959), and more recently for example by Götz (2013), Braun and Rosin (2015) and McDougall et al. (2015). The characteristics of disfluent behaviour have also been shown to be affected by non-linguistic factors such as gender and age (Bortfeld et al., 2001; Longauerová, 2016) or the type of context and the related level of anxiety or stress (Buchanan et al., 2014).

As regards the location of disfluencies within utterances, they frequently occur before long or complex constituents (Kjellmer, 2008; Watanabe et al., 2008), before grammatically complex constituents (Clark and Wasow, 1998), or before low-frequency words (Corley et al., 2007). Arnold et al. (2003) observe that they frequently precede items which are newly introduced into discourse. Biber et al. (1999) note that the location of the different types of disfluencies varies: unfilled pauses tend to separate major syntactic units, filled pauses lesser syntactic units and repeats may introduce any sentential constituent (e.g. a prepositional phrase). They also acknowledge that the location of disfluencies may be affected by cognitive problems resulting from the nature of the task and that cognitively demanding tasks may result in a higher variability in the type, frequency and location of hesitations. An interesting but not an entirely attested hypothesis is that hesitation phenomena are periodically distributed in spoken language production (Merlo & Barbosa, 2010).

The present study investigates the phenomenon of repeats, i.e. segments of speech which are involuntarily repeated in close proximity without adding any propositional content to the message. Along with filled pauses, repeats are amongst the most frequently occurring types of disfluency (Biber et al., 1999), which however need to be distinguished from repetitions, i.e. deliberate repetitions of words or phrases for rhetorical or other reasons. Example (1) is an illustration of a repeat, and whilst example (2) may be used as an illustration of a repetition of the intensifier *very* for added emphasis, it also shows that distinguishing between these two phenomena may be problematic: without access to the recording to judge the intonation we might not be able to determine whether the repetition of the word *very* is for reasons of emphasis or as a result of hesitation or planning difficulties.

(1) *I mean the the play is really great*

(2) *but the language really was very very nice*

In a seminal study, Clark and Wasow (1998) present repeats as analysable units composed of four subprocesses (initial commitment, suspension, hiatus, and restart). The speaker initially commits to a particular constituent, then suspends speech (for reasons of planning or other), he may fill the hiatus phase with a pause (filled or unfilled), and then resumes production by repeating from the start of the constituent. They observe that

the most frequently repeated words are those which are at the left-most side of the constituent: in English these positions are frequently occupied with function words rather than lexical ones. This is in line with both earlier (e.g. Maclay & Osgood, 1959) and later findings (Biber et al., 1999; Kjellmer, 2008) which show that the most typically repeated units are pronouns, articles, prepositions and contracted forms. Clark and Wasow (*ibid.*) claim that speakers produce repeats because they prefer to deliver continuous speech and therefore after the suspension of speech they start anew. Whilst this is a plausible hypothesis, it fails to explain why repeats are not produced by all speakers and after all points of speech suspension.

Within the context of non-native speech production, Lennon (1990) and Freed (2000) studied various aspects of fluency on a small sample of speakers in a study-abroad context. Whilst they do not provide a detailed analysis and typology of repeats, they observe changes in the frequency of disfluencies including repeats following the speakers' stay in an English-speaking country. Contrary to expectation, this change does not necessarily mean a decrease, which leads Freed to speculate whether the higher frequency of repeats may not be linked to the growing sophistication of the speakers' speech as a result of study abroad.

To date, the most thorough analysis of repeats used by non-native speakers is offered by Götz (2007, 2013). She compares German advanced learners of English with British native speakers and establishes patterns of overuse and underuse of different types of repeats based on Biber's et al. (1999) typology. These results are, however, hard to interpret as the studies do not describe in detail the methodological aspects of locating and classifying the repeats she was working with (the same is true of the above-mentioned studies by Lennon and Freed).

The current study aims to explore quantitative and qualitative aspects of the use of repeats by Czech advanced speakers of English and contribute to the ongoing discussion of the nature of disfluencies in non-native speaker spontaneous speech production. We are specifically interested in whether Czech advanced learners of English show any similarities in their use of repeats to those described in literature on native and non-native use of these disfluencies. This study thus extends what we view as a relatively underresearched area of L2 fluency and disfluency research.

2. Method

The data for the current study derives from the Czech subcorpus of the Louvain International Database of Spoken English Interlanguage (henceforth LINDSEI_CZ) (Gráf, 2017) which contains 50 approximately 15-minute recordings of advanced² English learners with Czech as their L1. This amounts to almost 13 hours of recorded material. The learners form a relatively homogeneous group of speakers of similar age (they were all 3rd- or 4th-year university students of English and American Studies), with 43 female and 7 male speakers. Such a homogeneous group does not allow for the exploration of age or gender related effects on fluency as were mentioned above. The orthographic

² LINDSEI uses institutional definition of advancedness (Gráf, 2015) and consequently the actual level of proficiency may not be the same for all of its speakers.

transcriptions of the recordings include disfluencies (filled and unfilled pauses, repeats, truncations and draws) which are counted as words. LINDSEI_CZ contains 123,761 words, of which 95,904 are words produced by the learners. The remaining 27,857 words uttered by the interviewer have not been included in the analysis.

To tag the instances of disfluencies I developed a simple interlinear, incremental tagging system (see Table 1 for examples). The first position of each tag contains the identification of the disfluency type (R = repeat, FS = false start, SC = self-correction). The second position is numerical and describes the length of the repeated phrase. Number 1 thus denotes a repetition of one word, number 2 of two words etc. The third position is numerical and expresses the number of times the phrase is repeated. The fourth position uses letters to encode the part of speech and various subtypes³. The fourth position is primarily used with repeats involving one word only. The fifth position is optional and helps distinguish subtypes (e.g. repetitions for rhetorical or discourse purposes).

Table 1. Examples of tags for repeats and their decoding

Example of a tag	Meaning of the tag
<R_1_3_P> I I I wouldn't do it myself	R = repeat, 1 = repeating one word, 3 = occurring 3 times, P = pronoun
which is . similar . (eh) <R_2_2> in many in many ways	R = repeat, 2 = repeating two words, 2 = occurring twice

In order to increase the reliability of the identification process I compiled a computer script for the automatic retrieval and tagging of repeated sequences. The script ignored any intervening pauses and fillers (and their combinations) so that sequences such as *I (erm) I* or *I . I* would still be identified as repeats. This follows Clark and Wasow's conception of repeats as analysable units, and more specifically the notion of hiatus, i.e. the space between the suspension and resumption of speech which may be left unfilled but may also be filled with different types of pauses.

Once all repeats were automatically tagged by the script, I listened to the individual files whilst following the tagged transcriptions to check whether the tagging was done correctly. This helped to distinguish between repeats and repetitions (usually disambiguated by intonation), and it also revealed instances in which the occurrence of two identical words next to each other were not cases of repeats. They were cases in which the co-occurring words were not part of the same constituent, as shown in examples (3–5), or sentence (the transcription does not use punctuation).

(3) *the Film Society have got it on on a Friday*

(4) *we went to see it it was Sunday morning*

(5) *we have had compliments from outside companies companies that normally deal with proper commercial cinemas*

³ Meaning of the codes in the fourth position of the tags: Ad – definite article; Ai – indefinite article; Ao – other determiner; B – preposition; C – conjunction; D – discourse marker; E – existential there; F – filler; G – adverb; Ip – infinitive particle; J – adjective; N – noun; O – other; P – pronoun; R – rhetorical; V – verb; W – wh-word; X – contraction

Only fully retraced elements were tagged as repeats, thus if the element involved any kind of rephrasing, it was tagged as a false start (FS), as shown in example (6). Also tagged as false starts were all instances in which only a part of the word was repeated as shown in example (7), and in example (8) in which each repetition of the initial syllable is tagged separately as a false start.

- (6) <FS_2> she didn't (eh) she . couldn't agree because
 (7) she was dissatisfied <FS_1> wi= with the painting
 (8) I mean <FS_1> av= <FS_1> av= . avoiding conflicts

As in similar studies (e.g. Maclay & Osgood, 1959; Clark & Wasow, 1998; Götz, 2013) graphic words are counted, which means that contractions are counted as one word (e.g. *I'm, it's* etc.). Biber et al. (1999) point out that this procedure is fully justifiable as contractions are processed as single items.

Once the tagging was completed, the files were analysed using AntConc (Anthony, 2014). Excluded from the count were all instances of repetitions for rhetorical (see example (2) above) or discourse purposes, as in example (9), and repetitions of filled pauses.

- (9) Interviewer: are you writing your bachelor's thesis at the moment
 Interviewee: not yet not yet I . plan to write it . during the[i:] Erasmus so

3. Results

A total of 2,311 sequences of repeated elements were identified. Once all instances of non-repeats as described in the preceding section were removed, 1,905 repeats remained for our analysis. As is shown in Table 2, more than three quarters (78.27%) of the bulk are formed by one-word repeats. Multi-word repeats are less common, with two-word repeats adding up to 19.3%, three-word repeats to 2.4% and longer repeats to approximately 0.1%.

Table 2. Frequencies of repeats of different length

Length of repeated segment	N	%
One word	1,498	78.3 %
Two words	369	19.3 %
Three words	45	2.4 %
Four words	2	0.1 %
More than four words	0	0 %
Total	1,914	100 %

Following Clark and Wasow's (1998) method, our discussion of repeats does not sub-categorize repeats with different types of hiatus or other variations. These are, however, relatively frequent: 20% of instances of one-word repeats include an unfilled pause (as in ex. 10), 5% include a filled pause (ex. 11), 3% include lengthening (ex. 12) and 3% include

an extension of a personal pronoun by a contracted form of a copular or auxiliary verb (ex. 13). The situation is similar for multi-word repeats.

- (10) *you can really enjoy <R_1_2_Ad> the . the view every morning*
 (11) *I'm a huge fan <R_1_2_B> of (erm) . of television series*
 (12) *<R_1_2_Ad> the: the lady seems to be pleased*
 (13) *I mean <R_1_2_P> I I've been doing that*

3.1 One-word single repeats

Table 3 shows a breakdown of the <R_1_2> type, when the speaker repeats a single word once. As is pointed by Biber et al. (1999: 1055) this is the most common type of repeat. In the present corpus, 1,349 such instances have been observed. The most commonly repeated elements are pronouns, conjunctions, prepositions, definite articles and contracted forms. These parts of speech also show a very high frequency across the board, as 98% of all of the speakers produced at least one instance of pronoun repetition, 90% of speakers repeat prepositions, 68% conjunctions, 68% contractions, and 66% repeat definite articles.

Table 3. Frequencies of one-word single repeats (tagged as <R_1_2_*>)

Repeated element	Count	%	Speakers involved
Pronoun	470	34.8 %	49 (98 %)
Preposition	164	12.2 %	45 (90 %)
Conjunction	163	12.1 %	34 (68 %)
Definite article	117	8.7 %	33 (66 %)
Contracted form	103	7.6 %	34 (68 %)
Adverb	61	4.5 %	28 (56 %)
Other types	59	4.4 %	32 (64 %)
Verb	53	3.9 %	23 (46 %)
Infinitive particle	42	3.1 %	24 (48 %)
Wh-word	39	2.9 %	24 (48 %)
Adjective	33	2.4 %	20 (40 %)
Noun	13	1.0 %	7 (14 %)
Indefinite article	12	0.9 %	10 (20 %)
Existential there	11	0.8 %	11 (22 %)
Other determiner	9	0.7 %	7 (14 %)
Total	1,349	100 %	

3.2 One-word multiple repeats

Multiple repeats of one word are considerably less common. The corpus contains 140 instances of triple repeats and 8 instances of quadruple repeats. As is shown in Table 4, these are again most frequently repeats of pronouns (40.5%), definite articles (8.8%),

conjunctions (8.1%) and prepositions (6.7%), but they occur in a much smaller selection of speakers: except pronouns which were repeated by 58% of speakers, all of the other types are repeated by fewer than 20% speakers.

Table 4. Frequencies of one-word multiple repeats (tagged as <R_1_3/4_*>)

Repeated element	Count	%	Speakers involved
Pronoun	60	40.3 %	29 (58 %)
Definite article	12	8.1 %	9 (18 %)
Conjunction	12	8.1 %	6 (12 %)
Preposition	10	6.7 %	9 (18 %)
Other types	12	8.1 %	8 (16 %)
Wh-words	8	5.4 %	5 (10 %)
Infinitive particle	6	4.0 %	4 (8 %)
Adjective	6	4.0 %	4 (8 %)
Contraction	6	4.0 %	6 (12 %)
Verb	4	2.7 %	4 (8 %)
Adverb	2	1.3 %	3 (6 %)
Indefinite article	2	1.3 %	1 (2 %)
Noun	1	0.7 %	1 (2 %)
Pronoun (repeated 4 times)	8	5.4 %	6 (12 %)
Total	149	100.0 %	

3.3 Multi-word repeats

Multi-word repeats detected in our corpus include 370 instances of two-word repeats and 45 instances of three-word repeats. The majority (241 instances, 65.14%) of our two-word repeats involve a subject followed by different types of complementation. As we can see in Table 5, the most frequent types are subject + copular verb (40.2%, see ex. 14), subject + auxiliary/modal verb (24.1%, see ex. 15), subject + lexical verb (12.9%, see ex. 16), and a combination of subject preceded by another word (17.43%), such as a conjunction as in ex. (17). Other instances are marginal.

- (14) <R_2_2> *it was it was just the inability to act*
 (15) <R_2_2> *I am I am planning my next visit to*
 (16) <R_2_2> *we see . we see children from the whole world*
 (17) <R_2_2> *when she when she actually sees the painting*

Two-word repeats frequently involve a verb (218 cases). These mostly (88.5%) include the combinations of subject + verb discussed above (see Table 5). Other combinations are rarer, such as verb + preposition (2.75%), copular/auxiliary/modal verb + lexical verb (1.4%), verb + object (2.3%), and to + infinitive (1.4%).

Table 5. Proportional distribution of repeats involving a subject and different types of complementation

	Copular	Auxiliary	Lexical	* + S	Existential	Adverb	Total
Subject	97	58	31	43	7	5	241
%	40.2 %	24.1 %	12.9 %	17.8 %	2.9 %	2.1 %	100 %

Table 6 displays the number of two-word repeats involving a preposition. The most frequently occurring repetitions of this type include the prepositions *in*, *on*, *to*, *of*, *with* and *as* (examples 18 and 19).

(18) *there are a lot of catchy phrases (erm) <R_2_2> in the in the play*

(19) *but I got to go <R_2_2> on a on a cruiseship*

Table 6. Proportional distribution of repeats involving prepositions + complement

in	on	to	of	with	as	for	from
14	8	7	7	6	4	4	4
23.0 %	13.1 %	11.5 %	11.5 %	9.8 %	6.6 %	6.6 %	6.6 %
at	like	about	through	by	instead of	Total	
2	1	1	1	1	1	61	
3.3 %	1.6 %	1.6 %	1.6 %	1.6 %	1.6 %	100 %	

Table 7 provides an overview of two-word repeats involving a conjunction. The most frequently occurring repetitions here include *wh*-words used as conjunctions (ex. 20), and then the conjunctions *and*, *so*, *that*, *as* and *if*. The majority of these repetitions (79%) are the combination of a conjunction followed by a subject (ex. 21), the remaining 21% are used within a nominal phrase (ex. 22).

(20) *it's massive and <R_2_2> when you when you really enter into it*

(21) *she's wearing (er) . a pretty dress <R_2_2> and he and he starts painting*

(22) *between the teacher <R_2_2> and the and the students*

Table 7. Proportional distribution of repeats involving conjunctions

	wh- word	and	so	that	as	if	but	after	before	than	Total
Conj. + wh	25	10	5	4	4	4	2	2	1	1	58
%	43.1%	17.2 %	8.6 %	6.9 %	6.9 %	6.9 %	3.4 %	3.4 %	1.7 %	1.7 %	100 %

With 45 instances, three-word repeats are considerably less frequent. Almost two thirds of them (64.4%) involve a subject and a verb (exs. 23–25), 9% are within a prepositional phrase (ex. 26), and then there are only singular instances of different types (e.g. ex. 27).

(23) *<R_3_2> it would be it would be very easy*

(24) *<R_3_2> it was an it was an awesome experience*

(25) *<R_3_2> I can see I can see the point*

(26) *<R_3_2> in the last in the last picture we can see*

(27) *<R_3_2> what would he what would he do*

3.4 Repeat rates

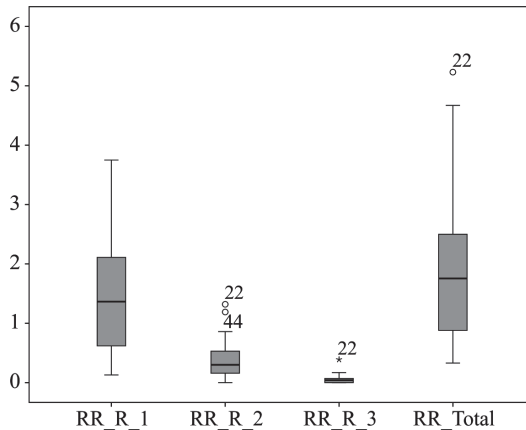
In the following section, I will inspect the frequency of occurrence of repeats as they were produced by the learners. Here a normalized frequency per hundred words (phw) is used, which I will henceforth refer to as “repeat rate”. Table 8 shows that the overall repeat rate in the whole corpus is 1.91 repeats phw (SD=1.18), which means that repeats occur once in every 52 words. One-word repeats occur at a rate of 1.47 phw (SD=0.94) (one instance every 68 words), two-word repeats at a rate of 0.37 repeats phw (SD=0.29) (once in every 270 words), and three-word repeats at 0.05 (SD=0.09) (once in every 2,000 words).

Table 8. Mean repeat rates per hundred words (phw)

Type of repeat	Repeat rate phw	SD
One-word	1.47	0.94
Two-word	0.37	0.29
Three-word	0.05	0.09
Total	1.91	1.18

The large standard deviations, however, indicate that there is a large inter-speaker variability in the production of repeats. Whilst the least disfluent speaker repeats at a rate of 0.33 repeats phw (one repeat every 303 words), the most disfluent one repeats at a rate of 5.23 repeats phw, producing one repeat every 19 words. Figure 1 provides a comparison of the repeat rates for the different types of repeats.

Figure 1. Ranges and the distribution of repeat rate values for one-word repeats (RR_R_1), two-word repeats (RR_R_2), three-word repeats (RR_R_3), and all repeats (RR_Total).



The values range from 0.32 to 5.12 repeats phw for one-word repeats, from 0 to 1.38 for two-word repeats, and from 0 to 0.39 for three-word repeats. All of the 50 speakers produced at least one instance of one-word repeat, 47 speakers produced two-word repeats, and 30 speakers produced three-word repeats.

4. Discussion

The purpose of this study was to investigate the use of repeats in the spontaneous spoken production of advanced learners of English. In particular, I investigated the types of repeated words and established a typology of repeats which occur in their spoken production. The most frequently occurring types of repeats in our corpus are one-word repeats. Within this group the most frequently repeated words are pronouns (including those with a contracted form with a verb), and especially personal pronouns. Taking into account the typical structure of English sentences, this finding confirms those of many previous studies which indicate that the use of repeats is a strategy connected with the planning of an utterance where most of the planning pressure is at the beginning of an utterance. Other major types of repeats contain articles or prepositions, which implies that planning pressure also increases and time needs to be gained at the beginning of noun or prepositional phrases. The last major category includes the use of conjunctions at the beginning of a clause, and also – if less frequently – within a noun phrase.

These results are in line with those of Biber et al. (1999)⁴, who, however, investigated the use of repeats in native-speaker English. It is interesting to observe that based on this comparison our advanced learners use similar strategies and with similar frequency as native speakers. The only comparable study exploring the phenomenon of advanced learner English is one carried out by Götz (2013), who in her own corpus of 50 advanced learners of English with L1 German observed the same types of repeats. The high dispersion in her data leads Götz (p. 109) to consider whether repeats are a fluency enhancing strategy which has been adopted only by the more advanced learners. However, she records a very similar dispersion in a parallel corpus of native speakers, which would rather seem to imply that different speakers might use different strategies to gain time for planning speech. Repeats are only one of these strategies, others including, for example, varying speech rate or the use of different pause, false-start or self-correction patterns. This area requires further investigation.

As regards multi-word repeats, the majority of them involve a subject and can thus again be found most frequently at the beginning of clauses. Our results cannot be compared here with other studies as the three main studies referred to above⁵ deal only with one-word repeats.

In agreement with Clark and Wasow (1998), Biber et al. (1999) and Götz (2007), our repeats are also accompanied by other types of disfluencies (in 15% of the cases), especially pauses or syllable lengthening. These occur either in the hiatus (i.e. between the repeated segments) but sometimes also before or after it. This illustrates that repeats themselves are not always sufficient means of gaining planning time and other strategies are adopted by the speakers in combination. Multiple-repetitions, i.e. those that involve more than two-fold repeats are fairly infrequent.

Whilst repeats were found to be used by all of our speakers, the large dispersion in their use shows that the group is rather heterogeneous and the strategy is not used by all speakers to the same degree. Further investigation is warranted especially with regard to

⁴ Biber et al. (1999) unfortunately do not report on the dispersion of their data.

⁵ i.e. Maclay and Osgood (1959), Biber et al. (1999), and Götz (2013)

their use of alternative choices of speech planning strategies, and this also raises a question whether the use of repeats is an area of pedagogical implications, and – more specifically – whether learners ought to be taught how to use repeats and fluency enhancing strategies in general.

5. Conclusion

This study has shown that as much as 2.5% of spontaneously produced speech in L2 learners is accounted for by the repetition of segments. This repetition might be seen either as a type of disfluency or as a fluency-enhancing strategy which allows the speaker to gain time for planning speech. The typology of these repeats has revealed that repeats are predominantly used at the beginnings of clauses or of nominal/prepositional phrases, where planning pressure is felt most acutely, and that the learners thus feel the need to plan not only at the beginning of clauses but also at the beginning of other constituents. More research is needed to explore the differences in the location of repeats produced by learners and native speakers.

However, not all of the learners appear to make use of this strategy and future studies of this matter should concentrate on finding which strategies are used as alternatives. Also, correlations can be sought between the use of repeats and proficiency, trying to determine whether more advanced learners use fluency enhancing strategies more effectively. Further research ought to be carried out investigating and explaining the similarities between the use of repeats in native and learner language. It would also seem worth our attention to see whether the use of repeats by L2 speakers mirrors their use of this strategy in their L1, and whether, indeed, this might be a specific area of language transfer. To this purpose, it would appear beneficial if learner corpora contained also samples of the participants' L1.

Previous studies of repeats in native speech show these to be a natural component of everyday speech. The present study shows that they are also frequent in L2 advanced speech. It is likely that the use of such time-gaining strategies positively affects fluency, and an important question must thus be raised whether L2 learners ought to consciously adopt such strategies and whether they can be helped in this process by explicit instruction.

ACKNOWLEDGEMENT

This research was supported by the Charles University project Progres 4, *Language in the shiftings of time, space, and culture*.

REFERENCES

- Anthony, L. (2014). *AntConc (Version 3.4.3)*. Tokyo: Waseda University. Retrieved from <http://www.laurenceanthony.net/>
- Arnold, J. E., Fagnanon, M. & Tanenhaus, M. K. (2003). Disfluencies Signal Thee, Um, New Information. *Journal of Psycholinguistic Research*, 32(1), 25–36.

- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow, England; New York: Pearson Education ESL.
- Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F. & Brennan, S. E. (2001). Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech*, 44(2), 123–147.
- Braun, A. & Rosin, A. (2015). On the Speaker-Specificity of Hesitation Markers. In: *Proceedings of the 18th International Congress of Phonetic Sciences. (ICPhS 2015)*. Glasgow: University of Glasgow.
- Buchanan, T., Laures-Gore, J. S. & Duff, M. C. (2014). Acute stress reduces speech fluency. *Biological Psychology*, 97, 60–66.
- Clark, H. H. (2002). Speaking in time. *Speech Communication*, 36(1), 5–13.
- Clark, H. H. & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, 37(3), 201–242.
- Corley, M., MacGregor, L. J. & Donaldson, D. I. (2007). It's the way you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, 105, 658–668.
- Fox Tree, J. E. & Clark, H. H. (1997). Pronouncing 'the' as 'thee' to signal problems in speaking. *Cognition*, 62(2), 151–167.
- Freed, B. F. (2000). Is fluency in the eyes (and ears) of the beholder? In: H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 243–265). Ann Arbor: University of Michigan Press.
- Götz, S. (2007). Performanzphänomene in gesprochenem Lernerenglisch: Eine korpusbasierte Pilotstudie. *Zeitschrift Für Fremdsprachenforschung*, 18(1), 67–84.
- Götz, S. (2013). *Fluency in Native and Nonnative English Speech*. Amsterdam; Philadelphia: John Benjamins Publishing Company.
- Gráf, T. (2015). Accuracy and fluency in the speech of the advanced learner of English (Unpublished PhD Thesis). Charles University, Prague, Czech Republic.
- Gráf, T. (2017). *Korpus LINDSEI_CZ*. Praha: FF UK.
- Kjellmer, G. (2008). Self-repetition in spoken English discourse. In: T. Nevalainen & I. Taavitsainen (Eds.), *The dynamics of linguistic variation: corpus evidence on English past and present; [... a selection of articles based on papers presented at the 27th Conference of the International Computer Archive of Modern and Medieval English (ICAME) in Helsinki in May 2006]*. Amsterdam: Benjamins.
- Lennon, P. (1990). Investigating Fluency in EFL: A Quantitative Approach*. *Language Learning*, 40(3), 387–417.
- Li, J. & Tilsen, S. (2015). Phonetic evidence for two types of disfluency. In: *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*. Glasgow: University of Glasgow.
- Longauerová, R. (2016). *Sociophonetic study of dysfluent behaviour in native English speakers* (Unpublished BA thesis). Charles University, Prague.
- MacGregor, L. J., Corley, M. & Donaldson, D. I. (2009). Not all disfluencies are equal: The effects of disfluent repetitions on language comprehension. *Brain and Language*, 111(1), 36–45.
- Maclay, H. & Osgood, C. E. (1959). Hesitation Phenomena in Spontaneous English Speech. *Word*, 15(1), 19–44.
- McDougall, K., Duckworth, M. & Hudson, T. (2015). Individual and Group Variation Dysfluency Features: A Cross-Accent Investigation. In: *Proceedings of the 18th International Congress of Phonetic Sciences. (ICPhS 2015)*. Glasgow: University of Glasgow.
- Merlo, S. & Barbosa, P. A. (2010). Hesitation phenomena: a dynamical perspective. *Cognitive Processing*, 11(3), 251–261.
- Rühlemann, C. (2006). Coming to terms with conversational grammar: 'Dislocation' and 'dysfluency'. *International Journal of Corpus Linguistics*, 11(4), 385–409.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.
- Skehan, P. (2003). Task-based instruction. *Language Teaching*, 36(1), 1–14.
- Voss, B. (1979). Hesitation phenomena as sources of perceptual errors for non-native speakers. *Language and Speech*, 22(2), 129–144.
- Watanabe, M., Hirose, K., Den, Y. & Minematsu, N. (2008). Filled pauses as cues to the complexity of the upcoming phrases of native and non-native listeners. *Speech Communication*, 50, 81–94.

RESUMÉ

Spontánní mluvený projev je charakteristický mj. tím, že se v něm vyskytuje velká řada tzv. disfluencí, např. opakování slov, užívání pauz, falešných začátků či přeformulování částí promluv. Jde o jevy, které mohou být chápány jako prvky narušující plynulý tok řeči, ale zároveň mohou být považovány za strategie, které plynulý tok řeči podporují, a to především tím, že mluvčímu umožňují získat čas na plánování promluvy.

Předložená studie se zabývá výzkumem jednoho z nejčastějších typů disfluencí, a to opakování slov. Na základě mluveného žákovského korpusu pokročilé angličtiny mluvčích s češtinou jako mateřským jazykem zkoumá spontánní projev (o celkové délce téměř 13 hodin) 50 studentů anglistiky na FF UK. V tomto vzorku bylo identifikováno 1905 případů opakování. Téměř 80 % z nich tvoří opakování jednoho slova, téměř 20 % opakování dvouslovných segmentů a pouze asi 2 % opakování segmentů delších.

Kvalitativní analýzou bylo zjištěno, že opakování nejčastěji obsahují osobní zájmena, spojky a předložky, což vypovídá o tom, že mluvčí této strategie využívají na začátku vět a větných frází, kde je pocíťován největší tlak na plánování promluvy. Tyto výsledky jsou srovnatelné se studii projev rodilých mluvčích i mluvčích angličtiny jako cizího jazyka s jiným mateřským jazykem (němčinou).

Pokud vnímáme opakování jako účinnou strategii při řečovém managementu, nabízí se otázka, zdali tyto strategie nemají být součástí výuky cizích jazyků.

Tomáš Gráf

Department of English Language and ELT Methodology

Faculty of Arts, Charles University

tomas.graf@ff.cuni.cz