UDC 811.111:811.163.41
https://doi.org/10.18485/ms_zmsfil.2022.65.2.7
Изворни научни рад

Tatjana Paunović

TEMPORAL PROPERTIES OF SPEECH AND THE PERCEPTION OF READING (DIS)FLUENCY: SERBIAN EFL VS. L1 SERBIAN AND L1 ENGLISH*

This paper presents a corpus-based study of the temporal properties of read speech in EFL, compared to L1 English and L1 Serbian, and the perception of its fluency by two groups of listeners –students and teachers. The analysis included acoustic measurements of variables, grouped together as speech-rate, pause, and disfluency variables. The listeners rated all the speech samples on a 5-point likert scale. The findings showed that EFL differed from L1 English by lower speech-rate variables, fewer pauses overall, but more within-IU pauses and disfluences. It showed no mother-tongue influence for speech-rate, but did for pauses and the mean length of run. Some temporal properties were significant predictors of (dis)fluency. The L1 Serbian corpus revealed some dialect-specific temporal properties.

Key words: fluency perception, speech rate, pauses, disfluencies, read speech.

У овом раду представљено је истраживање темпоралних својстава говора у корпусу текстова које читају студенти енглеског као страног језика (EFL корпус), у поређењу са изворним говорницима енглеског (L1-E) и српског језика (L1-S), као и перцепције течности говора од стране ученика и наставника. Анализа је обухватала акустичка мерења варијабли груписаних у категорије: брзина говорења, паузе и „поштапалице" (вокални елементи који испуњаваку паузе). Слушаоци су оцењивали течност свих узорака говора на петостепеној ликертовој скали. Резултати су показали да се EFL корпус разликује од L1-E корпуса по нижим вредностима брзине говорења и пауза, али уз бројније паузе унутар интонацијских целина, као и знатно бројније поштапалице. Није утврђен утицај матерњег језика код брзине говорења, али јесте код пауза. Утврђена је значајна повезаност више темпоралних својстава и оцене течности говора. Нека својства у L1-S корпусу могу се повезати са урбаним нишким варијететом призренско-јужноморавског дијалекта.

Кључне речи: перцепција течности говора, брзина говорења, паузе, поштапалице, читање као говорна вештина.

1. Introduction.[1] In language learning, fluency is a concept commonly considered central to oral performance, and closely related to the speaker's overall language proficiency (cf. De Jong et al. 2012a; b). However, despite abundant and varied empirical research, what is perceived as fluent speech in either first (L1) or second or foreign language (L2) is still not clearly specified. The understanding of the concept has evolved over the past decades, to a broader view that fluency is not "flawless fluidity" (Crible 2018: 2), but rather a skilful and efficient use of various communicative devices for a range of communicative purposes. This kind

* This study was part of the project (No. 455/1-1-1-01) *English Studies in the Digital Age,* conducted at the University of Niš, Faculty of Philosophy, Serbia.

[1] Some of the results of this research study were presented at the Fifth Belgrade International Meeting of English Phoneticians, BIMEP 2020 online, 30-31 March, Faculty of Philology, University of Belgrade.

of view, that some disfluencies in speech, particularly self-corrections, formulaic language and some pause-fillers, can be not only potential "symptoms" of problems but also hearer-oriented "signals" deliberately used for some communicative purposes (Clark – Fox Tree 2002: 75), is endorsed by many contemporary fluency researchers. Even so, when it comes to different modalities of speech, while in unscripted speech and spoken interaction various communicative devices may be relevant predictors of fluency perception – or fluency in the "broad sense" (Lennon 1990) – in read or scripted speech the temporal properties such as speech rate and pause frequency, distribution, and duration – or fluency in the "narrow sense" (Lennon 1990) – seem to be central to the listener's perception of the speaker's fluency.

This paper focuses on fluency in the narrow sense (Lennon 1990), in order to observe the temporal properties of read speech in L1 Serbian, L1 English, and Serbian learners' L2 English i.e., English as a foreign language (EFL). The study aimed to explore the relatedness of the temporal properties of speech to the listeners' evaluation of fluency, and to compare the temporal properties in L1 and L2. Read speech is chosen over spontaneous or conversational speech for two reasons. Firstly, the fluency-related temporal properties of Serbian L2 English have not been investigated yet, in either scripted or spontaneous speech. Additionally, while some research studies with different L2 English speakers have shown that the temporal properties (pauses, speech rate, hesitations) contribute to higher fluency in scripted speech compared to unscripted speech (e.g. Eren et al. 2022), others found a more complex picture of fluency and disfluency markers across speech genres (scripted vs. spontaneous conversation) and L2 proficiency levels (e.g. Kosmala – Crible 2022), i.e. that different speaking styles often reflect different temporal correlates of fluency (Gut 2009: 87–88), and that certain phenomena are "style- or 'habitat'-specific" (Wagner et al. 2015: 10). Therefore, exploring a wide range of temporal properties in the more controlled context of read speech would represent a solid comparison ground for further investigations of the perception of (dis)fluency in other speech modalities.

1.1. Fluency and the temporal structure of speech. Research on both L1 and L2 fluency builds on the view that observable properties of speech reflect the underlying speech-generating processes, cognitive as well as performative. From the earliest studies by Goldman-Eisler (1961a; b; c; 1968), the temporal structure of speech has been observed as related to the "social, emotional, cognitive" conditions of the context (Goldman-Eisler 1961b: 232), as under certain conditions the hesitation phenomena of filled and unfilled pauses, may "reflect different internal processes" (Goldman-Eisler 1961a: 18). Fillmore (1979) proposed four main aspects of fluency, which, beside speaking without many pauses and without many fillers (e.g., *for example*, *you know*, etc.), also involved speaking appropriately to the communicative social context, and showing a creative use of the language. Reaching back to Fillmore (1979), Segalowitz describes fluency as a "property of L2 use that emerges from the complex interplay of many factors interlinked in a dynamical system" (Segalowitz 2010: 28) and proposes a model of fluency comprising *cognitive fluency* (processing speed, lexical access, the use of linguistic resources for sociolinguistic and psychosocial functions, etc.); *utterance fluency* (speech rate, pauses, hesitation phenomena), as well as, beside motivation to communicate,

the features of the interactive and communicative social context that affect the *perceived fluency* (Segalowitz 2010; 2016).

Since the perception of fluency is complex and affected by various factors, the relationship between the temporal properties of speech and the perception of fluency is not yet completely clear. Segalowitz states that all types of fluency involve fluidity or flow, depend on the appropriate timing of speech, and are based on a temporal flow in the use of language (Segalowitz 2010: 4). Still, he points out Lennon's (1990) observation that increased fluency is related to speech rate and pause-time in a complex way, as with increased fluency the participants in Lennon's study produced fewer silences (rather than faster talk), but also showed an *in*creased and not *de*creased ability to self-correct (Segalowitz 2010: 32). Along similar lines, in her corpus-based study of the properties of L2 English and German, Gut (2009) defines fluency as a concept that comprises a continuum, ranging from underlying processes and competences to individual production processes and measurable temporal dimensions of speech, such as speech rate, length of runs, or disfluency markers (repetitions, repairs, pauses) (Gut 2009: 78–79).

Similarly, in her corpus-based study of fluency in L1 and L2 English, Götz (2013), too, relies on Lennon's "two key fluency markers": speech-pause relationships, and the frequency of dysfluency markers, such as filled pauses and repetitions (but not necessarily self-corrections) (Lennon 1990: 388 in Götz 2013: 13). Götz distinguishes between "primary fluency variables", i.e., those features that always occur in each speaker's speech, such as temporal variables (speech rate, unfilled pauses, or phonation – time ratio), and "secondary" variables of fluency that do not always occur (e.g., discourse markers) (Götz 2013: 8). Also, Crible (2018) views fluent and disfluent as a "scale or a continuum rather than clear-cut categories" (Crible 2018: 4) and concludes that each element (or cluster of elements) needs to be analysed in a particular local context in order to "diagnose" whether it is a symptom of disfluency or a deliberate communicative signal (Crible 2018: 3). More specifically, in their corpus study, Kosmala and Crible (2022) found two distinct patterns of distribution for "filled pauses" (*euh* and *eum*) – the initial position, commonly with a discourse marker, and the medial position, accompanied by other hesitation markers, and that these two reflect different, "fluent" and "disfluent," uses of filled pauses.

Therefore, it can be said that abundant research on speech fluency has broadened the notion, but it has also shown that some temporal variables, particularly speech rate and silent pauses, are immediately relevant for the perception of fluency. Although many researchers object to the lack of agreement on the exact measures of temporal structure (cf. Gut 2009; Kormos 2006; Segalowitz 2010),[2] others point out that research findings suggest "strong associations between utterance fluency and perceived fluency" (e.g., De Jong et al. 2012b: 896).

Several temporal measures are widely recognized as potentially relevant in this respect:

[2] For instance, whether to observe speech rate in words per minute, syllables per minute, syllables per second, or even phones per second, or what threshold to define as the minimum cut-off point for silent pause duration – which in different studies ranges "anywhere between 100 ms and 1,000 ms" (De Jong – Bosker 2013: 17), sometimes even from 60 ms (Kendall 2013) up to 3,000 ms (cf. the studies listed in Gut 2009: 80).

1) *Speaking rate* – the total number of syllables produced divided by the total speaking time, including pauses.

2) *Articulation rate* – the total number of syllables produced divided by the speaking time excluding silent pauses (*articulation time, phonation time*). Although articulation rate was traditionally regarded as more "personally constant" (Goldman-Eisler 1961c: 171), it is a relevant parameter in L2 speech, since, as pointed out by Redford, "[r]apid, stable, and efficient execution of complex movement sequences requires extensive practice, and […] can continue to improve over many years" (Redford 2015: 388).

3) *Phonation–time ratio* – the ratio of actual speaking time excluding pauses to the total speaking time with pauses included (i.e., articulation time divided by the speaking time).

4) With respect to *pauses*, their overall *number*, *duration*, and *frequency* are commonly measured, both for *unfilled* or *silent pauses*, defined as "silence or the occurrence of non-speech acoustic events such as breathing and noise" (Gut 2009: 80), and for *filled pauses*, described as including sound fillers (*er*, *uh*, *um*, *erm*) and sound prolongations (drawls). Other types of disfluencies (*repetitions*, *echoes*, *restarts*, *repairs*, *mispronunciations*), and the use of "formulaic devices" or "small words" (*well*, *right*, *okay*, *you see*, cf. Gut 2009: 82) are also included in some studies, and sometimes *the degree of hesitancy* (DOH) is also calculated, as the number of filled pauses divided by the total number of pauses, times 100 (cf. Tomić 2017). The *pause ratio* (total pause time as the percentage of the total speaking time) is also a common measure in a number of studies.

Furthermore, some researchers propose calculating pause frequency as a function of language produced (e.g., per 100 syllables), others as the number of pauses per unit of time (e.g., minute) (cf. Gut 2009: 80). This is especially relevant for investigating developmental phenomena in L1 and L2 acquisition, since younger L1 and less proficient L2 speakers may produce less speech over time, as well as shorter utterances (Redford 2013; 2015: 387).

5) *Mean length of run* – the average number of syllables between pauses.

6) Some additional measures aim to capture the relatedness of the temporal properties of speech to the structural context. For instance, Segalowitz (2010: 39, based on Kormos 2006: 162) lists: *PACE* (stressed words/ minute) – the number of stressed syllables in a time unit; and *SPACE* (stressed words/ total words) – the ratio or proportion of stressed words to the total number of words.

1.2. Research on the temporal properties and perception of fluency. Of the extensive previous research, only some findings most directly relevant for this study will be summarized, those pertaining to L2 English, read speech, or the temporal properties of Serbian.

Gut's (2009) corpus analysis of 161 L2 English speech recordings elicited in three speaking styles (reading, retelling, and free speech) showed that the silent pause ratio was the only measurement that did not vary with the speaking style. In the L2 English corpus as a whole, the participants produced: 9.75% silent pauses and 15.1% filled pauses, the speech rate of 2.63 σ/sec (syllables per second) and articulation rate of 4.19 σ/sec; the mean length of run was 6.22 σ (Gut 2009: 94). Reading, however, was characterised by a higher speech rate and articulation rate

than both free speech and retelling, as well as a longer mean length of run, 7.3 σ, compared to 5.8 σ in free speech and 5.4 σ in retellings. Very few filled pauses were produced in reading, 1.5%, compared to 15.3% and 23.46% in the other two styles respectively (Gut 2009: 98). Compared to L1 English speakers, L2 English speakers' reading had slightly higher silent pause ratio (9.8% vs. 9.03%); they also had a slower speech rate (3.3. vs. 4.1 σ/sec) and articulation rate (4.4 vs. 5.6 σ/sec) and a shorter mean length of run (7.3 vs. 9.44 σ). Most notably, although the filled pause ratio in L2 English was not very high (1.5%), filled pauses were virtually absent from L1 English reading (Gut 2009: 99). Therefore, Gut concludes that the most relevant indicators of L2 fluency are the articulation rate and the mean length of run (Gut 2009: 111).

In her study of pauses in L1 and L2 speech, De Jong (2016) found, in line with similar previous research (e.g., Tavakoli 2011), that L2 speakers paused more often than L1 speakers within utterances, and with longer pause duration. The logistic regression analyses of L2 Dutch produced by 72 participants of different L1 backgrounds, showed that that L1 and L2 speakers do not differ in pause production at utterance boundaries. However, within utterances, L2 speakers were more likely to pause, either with a filled or with a silent pause, but more proficient learners produced fewer such pauses (De Jong 2016: 129). De Jong and Mora (2019) also found that L2 speech had more silent pauses and a slower articulation rate, but that the duration of silent pauses was not different in L1 and L2 speech (De Jong – Mora 2019: 237).

Focusing on the listeners' perception of fluency, Derwing and colleagues (2004) investigated untrained listeners' assessments of fluency as related to the temporal and hesitation features of Mandarin L2 English. A high parallelism was observed between the raters' judgment of fluency and the temporal measures of pausing and speech rate (standardized pruned syllables, with excluded hesitations). Therefore, the authors conclude that temporal measures, albeit not the only ones, are an important indicator of fluency, closely related to the listeners' assessment of fluency (Derwing et al. 2004: 672).

The study by Comeaux and Thomson (2019) included L2 English speech samples produced by 10 L1 Mandarin and 10 L1 Slavic speakers in a picture description task; matched versions of these samples were produced so as to contain a) no hesitations, b) hesitation markers at clause boundaries (*um*, *uh*, or silence), or c) hesitation markers within clauses. The findings showed that hesitation-free samples were rated the highest on fluency, and that speech with unfilled pauses at clause boundaries was rated relatively high; however, the samples containing filled pauses were rated more negatively than samples containing unfilled pauses (Comeaux – Thomson 2019: 110), as listeners preferred speech samples with unfilled pauses and rated them more favourably in 70% of cases (Comeaux – Thomson 2019: 111).

Similarly, Kahng (2018) observed the effect of pause location on the perception of L2 fluency in 31 Korean EFL learners, focusing on the frequency, length, and distribution of silent pauses in one experiment, and on the pause location in the other experiment involving spontaneous speech. The findings showed that the location of the pause played a significant role in the perception of both L1 and L2 fluency, and Kahng concludes that listeners seem to connect clause-internal pauses

with the speakers' lower cognitive fluency. These findings fall in line with the findings of Shea and Leonard (2019) that pause measures are related to the learners' proficiency scores: longer pauses are better predictors of L2 proficiency, unfilled pauses are a strong predictor of L2 proficiency, but particularly mid-clause pauses. Similarly, analysing a large corpus produced by EFL learners and English as a second language users, Götz (2019) found that the number of filled pauses was significantly different and indicative of the speakers' proficiency levels, but also of context variables, such as the L1 background, and the L2 learning onset age.

Finally, Segalowitz (2010: 167) points out that it is of particular importance to compare a speaker's L2 temporal properties to the baseline properties of the speaker's L1 speech. Similarly, De Jong and Mora (2019: 228) found that individual differences in L2 fluency can be accounted for partly by the differences in the speaker's L2 proficiency, and partly by personal ways of speaking that surface in both L1 and L2 speech. They conclude that there are sets of speech features that identify a speakers' personal speaking "style" (De Jong – Mora 2019: 229) and that L1 and L2 measures of fluency (at least in spontaneous speech) are strongly related (De Jong – Mora 2019: 236).

Unfortunately, there are only a few studies dealing with temporal properties of speech either in L1 Serbian or in Serbian L2 English. Ilse Lehiste (2000) compared oral readings of a children's poem (four stanzas, four lines each) by four L1 Serbian speakers. The measurements included the duration of each line, the pauses between the lines, the metric feet, and the duration of the stressed syllable nucleus in the first syllable of the metric foot (Lehiste 2000: 590). The results showed that the speakers distinguished the internal structure of the stanzas in terms of pause duration – the longest pause was produced at the end of the stanza (average 857 ms), while the pause after the second line was longer (average 508 ms) than after the first and third lines (averages 323 ms and 319 ms).

Marković, Jakovljević, Milićev, and Milievć (2015) investigated the perception of synthesized vs. natural speech in L1 Serbian, particularly the role of pauses. In several perception tasks (quality assessment, intelligibility, comprehension) based on texts in Serbian produced by the AlfaNum speech synthesizer and by a professional actor, L1 Serbian speaker, as well as on synthesized speech with modified pauses, the results showed no differences between natural and synthesized speech in terms of perceived intelligibility, but in comprehensibility tasks the performance was better for natural than for synthesized speech, and the characteristics of pauses were central to the listeners' preference for natural over synthesized speech. The authors point out that the modified prosody in synthesized speech, i.e., inserting pauses between intonational phrases, contributed to better performance.

Lastly, in a recent forensic study, Tomić (2017) focused on the temporal parameters of spontaneous speech (articulation rate, speaking rate, degree of hesitancy, percentage of pauses, average pause duration), in the speech samples produced by ten female speakers in their L1 Serbian and L2 English. The findings showed that the mean articulation rate for Serbian was 11.63 phones per second (SD=0.82), while for English it was 10.52 phones per second (SD=0.48). Also, the mean speaking rate for Serbian was 8.07 phones per second (SD=0.87) and for

English it was 7.05 (SD =0.59) (Tomić 2017: 133). The mean degree of hesitancy was lower for the participants' L1 Serbian (41%, SD=13.86) than for their L2 English (49.37%, SD=16.41), though with large between-speaker variations in L2 English (Tomić 2017: 135).

2. Aims, methods, and procedures. The present study is a corpus-based exploration of the temporal features of speech produced in read-aloud tasks by proficient (CEFR B2+ level) Serbian EFL students. The aim was to observe the participants' L2 temporal properties against the background of L1 Serbian and L1 English read speech produced by speakers of comparable age and education. In the production part of the study, the analysis focused particularly on two aspects of temporal organization: the speech rate (speaking rate and articulation rate), and the frequency, duration and distribution of pauses, silent and filled. Secondly, the aim was to explore the relatedness of these temporal properties to the listeners' evaluation of the speakers' fluency, so in the perception part of the study the speech samples were rated for fluency by two different groups of listeners: untrained students and EFL teachers.

2.1. Production corpora. The main corpus of read speech (EFL corpus) was collected by recording 19 EFL students (9 male, 10 female, aged 19–22, mean 20.4) reading 11 short stories (270 to 420-word anecdotes). The reference corpus of L1 Southern British English (*L1-E corpus*) was produced by 4 students (male, aged 19–22) reading a 240-word story (anecdote) selected from the reading materials used for the main corpus. The reference corpus in L1 (Southern) Serbian (*L1-S corpus*) was produced by 4 students (male, aged 21–22) reading a 260-word story (anecdote) in Serbian, selected randomly from the participants who contributed to the main corpus. It should be noted that all of them were born and raised in Niš, i.e., that their mother-tongue variety of Serbian was the urban Niš variety of the Prizren-Južna Morava dialect. All the reading materials contained both narrative paragraphs and dialogues, and were closely matched for syntactic and lexical complexity, but not for intonational phrase weight, i.e., the number of syllables, which can affect the duration of silent pauses (Krivokapić 2007; 2012), so the influence of this factor remains an issue for further research.

2.1.1. Production corpora analyses. The recordings were transcribed by aligning each participant's reading with the original story text and hand-correcting the transcription to match the speaker's actual production, i.e., to include omissions, wrong words, filled pauses, and disfluencies, i.e., "major breaks in the speech flow […] leading to some sort of disruption" (Crible 2018: 9). All such disfluency elements were manually identified, included in the transcription, and coded for their type, relying on the adapted classification by Crible et al. (2016, in Crible 2018: 22-23), as:

1) filled pauses – both E and S: *er*, *uhm*, *uh*, *eh*, *euh*;

2) false starts – E: [*They come*] *They came*...; S: [*Tamo*] *Taman sam bila*…;

3) truncation – E: *bor-* [(break) *er*] *groceries*...; S*: po-* [(break) *uh*] *prostrla*…;

4) repetition – E: *and bought* [(break) *uh*] *and bought*...; S: *kao što su* [(break) *er*] *kao što su*…;

5) editing fillers – E: *oops, I'm sorry,*...; S: *mislim*,…;

6) misarticulations that involved an observable prolongation of a consonant or a vowel (*drawl*) which could be interpreted as a 'stalling' technique or a hesitation-filler, were transcribed and coded as another element of disfluency, e.g., E: *all her*[*rrrrr*] *deceased friends*; S: [*nnn*]*naokolo*. Other mispronunciations were disregarded if they were not disruptive to the speech flow, and so were segmental prolongations that occurred at the end of an utterance or intonation unit (IU).

The syllable counts of the transcripts were performed by an on-line automated word- and syllable counter (*Word Count*), and then, as they were derived from the orthographic transcripts, hand-corrected where necessary, relying on the auditory and visual inspection of the recordings.[3] However, the corpus was not phonetically transcribed, and the analysis did not take into account possible syllable reductions (e.g., *family* produced as 2 syllables), i.e. it was based on the number of *intended* (or *canonical*) syllables, not *realised* syllables (cf. Koreman 2006; Trouvain et al. 2001). This method was chosen over the acoustic method (identifying the peaks in the acoustic signal) or the auditory method (identifying the realized/ perceived syllables), relying on previous corpus-based research (e.g., Kendall 2013: 62).

Since the study focused on read and not spontaneous speech, the minimum duration threshold for silent pauses was not set in advance. Instead, in each participant's speech, all the silent intervals were identified and coded for their structural and prosodic position based on the original text, as 1) between-utterance breaks (major pauses), 2) breaks between intonation units within an utterance (minor pauses), 3) breaks within intonation unit (hesitations), and 4) segment-related silences (stop or affricate closure). This information was coded as a categorical variable, and pauses were defined as silences from categories 1 to 3. The same position coding was applied for filled pauses and disfluencies. The non-vocal events (audible breathing, sighs, lip-smacks, coughs) were treated as unfilled pauses. The duration of all these events was measured in seconds (sec) (Praat v. 5.2.03, Boersma – Weenink 1992-2010).

The acoustic measurements taken for each speaker included the following:

1. Total speaking time; 2. The number and duration of unfilled pauses; 3. The number and duration of disfluencies; 4. The length of run (number of syllables between pauses); 5. The duration of each run.

From these, the following derived measures were calculated:

6. Speaking rate – total number of syllables produced divided by the total speaking time including pauses (σ/sec); 7. Articulation (phonation) time – speaking time without unfilled pauses; 8. Articulation rate – total number of syllables produced divided by the articulation time excluding unfilled pauses (σ/sec); 9. Phonation / time ratio – articulation (phonation) time divided by the total speaking time; 10. Pause frequency as the number of pauses per 100 syllables; 11. Pause frequency as the number of pauses per minute.

The statistical procedures applied included means comparisons, correlations, regression, analysis of variance, and two-way between-groups analysis of variance

[3] For instance, where the final orthographic *–e* was identified as a syllable nucleus, as in *late*, or *O.K.* treated as a one-syllable word, etc.

to compare EFL to L1-E and L1-S. Non-parametric procedures were chosen after normality testing.

2.2. Perception of fluency. The recordings were evaluated for fluency by two groups of raters: 1) naïve listeners, 35 first-year students with no prior training in either phonetics or teaching, and 2) three TEFL teachers, one with over 20 years of experience and two novice teachers. The listeners were asked to rate each speaker for being fluent on a five-point scale, from 1= not at all to 5= highly fluent. These two evaluations, the students' and the teachers', were treated as separate variables in statistical analyses. No definition of the term 'fluency' was provided prior to listening, and no further explanations. De Jong and colleagues (2012b: 897) warn that if no instructions are given, raters use their own definition of fluency i.e., fluency is understood as a listener construct, which can be problematic in oral proficiency assessment. Still, we opted for this methodological choice since linking temporal properties of speech to speaker's proficiency level was not the aim of this study. Focusing the listeners' attention to some temporal properties specifically, as was useful in some previous research (Derwing et al. 2004; Rossiter 2009) might skew the possible relatedness of the temporal properties and fluency ratings, which we wanted to establish independently (Wagner et al. 2015: 10), in order to observe possible differences between the two groups of raters.

3. Results and discussion

3.1. EFL corpus – Temporal properties. The EFL corpus was characterized by a large variation between speakers, as indicated by the variability measures, shown in *Table 1* together with the means for the relevant variables, grouped into three thematic categories: *a)* speech rate, *b)* unfilled pauses, and *c)* filled pauses and disfluencies.

The speech rate variables in our corpus show similar results to some previous studies, e.g., in L2 English reading in Gut's (2009) study (articulation rate 4.4 σ/sec; speech rate 3.3 σ/sec). However, pause variables differ more, e.g., the silent-pause ratio was considerably higher in our EFL corpus than in Gut's L2 corpus (9.81%). The disfluency ratio is also considerably higher than only 1.35% reported in Gut's study (Gut 2009: 99).

Table 1a. *EFL corpus*: Means and variability measures for the *speech-rate* variables.

	Mean	Std. Dev.	Variance	Skewness		Kurtosis	
				Statistic	Std. Error	Statistic	Std. Error
Speaking RATE	3.630	.205	.042	-.295	.524	.196	1.014
Speaking rate PRUNED	3.740	.196	.039	-.535	.524	.443	1.014
Articulation rate	4.529	.328	.108	-.223	.524	.275	1.014
Phonation–time ratio (AT/ ST)	80.322	4.045	16.364	.372	.524	.443	1.014
PACE (stressed syllables / min)	94.325	6.806	46.320	-.501	.524	.224	1.014
SPACE (stressed / all words)	.575	.051	.003	3.111	.524	11.744	1.014

Table 1b. *EFL corpus*: Means and variability measures for the *pause* variables.

	Mean	Std. Dev.	Vari-ance	Skewness		Kurtosis	
				Statistic	Std. Error	Statistic	Std. Error
Number of pauses	46.74	12.736	162.205	.485	.524	-.043	1.014
Pauses – total duration	19.937	6.403	41.001	-.027	.524	-.727	1.014
Pause RATIO (% of sp. time)	19.661	4.051	16.416	-.358	.524	.411	1.014
Pauses per 100 syllables	12.851	2.165	4.685	.535	.524	-.395	1.014
Pauses per minute	27.940	4.688	21.976	.579	.524	-.483	1.014
No of pauses within IU	3.84	3.202	10.251	.262	.524	-1.540	1.014
Duration of pauses within IU	.772	.993	.986	1.734	.524	2.933	1.014

Table 1c. *EFL corpus*: Means and variability measures for the *disfluency* variables.

	Mean	Std. Dev.	Vari-ance	Skewness		Kurtosis	
				Statistic	Std. Error	Statistic	Std. Error
Number of disfluencies	8.68	5.578	31.117	.310	.524	-1.006	1.014
Disfluencies – total duration	3.065	1.950	3.841	.793	.524	1.279	1.014
Disfluencies % of sp. time	2.964	1.653	2.731	.501	.524	.951	1.014
Disfluencies per 100 syllables	2.362	1.537	2.361	.433	.524	-.575	1.014
Disfluencies per minute	5.049	3.207	10.286	.403	.524	-.628	1.014
Degree of hesitancy (DOH)	18.869	11.560	133.643	.245	.524	-.984	1.014

The disfluencies identified in the EFL corpus were not particularly numerous, as 12.6% of the total number of IUs in the corpus contained a disfluency, but they were of various types. Of the total of 165 disfluencies, false starts and repetitions were the most frequent, followed by the [*er*] pause filler and consonant prolongations, as shown in *Table 2*.

The mean length of run in the EFL corpus was 5.564 σ (syllables) per IU (SD=.621; variance = .386), with the range of 2.416 (min.= 4.565; max.= 6.981). The mean duration of runs was 1.522 sec (SD=.221; variance=.049), with the range of .782 sec (min.= 1.166; max.= 1.948).

Table 2. *EFL corpus*: The frequencies of the *disfluency types*.

Total IU in EFL corpus = 1305	Frequency	Valid %	Cumulative %
explicit comment	0	.0	.0
other vocalizations	**15**	1.2	1.2
false start or repetition	**66**	5.1	6.2
prolonged consonant	**26**	2.0	8.2
prolonged vowel	**9**	.7	8.9
[um] pause filler	3	.2	9.1
[er] pause filler	**46**	3.5	**12.6**
∑	**=165**	(12.6%)	
No disfluencies	1140	87.4	100.0

3.2. L1-E CORPUS – TEMPORAL PROPERTIES. Being smaller, the L1 English reference corpus showed less speaker variability, particularly for pauses and disfluencies. *Table 3* sums up the means and variability measures for all the relevant variables grouped into three categories, as for the EFL corpus. The small total number of disfluences in the L1-E corpus was expected, but the number of disfluencies was also proportionally much smaller (5.5%, compared to 12.6% in the EFL corpus). Still, almost all disfluency types were observed, even if the corpus was smaller, as shown in *Table 4*. Also, it should be noted that in some previous research, filled pauses were completely absent from L1 English read speech (e.g., Gut 2009: 99).

The mean length of run in L1 English was 5.268 σ/IU (SD=.164, variance=.027), with the duration of 1.293 sec (SD=.105, variance=.011). In this respect, our L1 English findings differ from Kendall's (2013), who found that the median number of σ/utterance for the reading data was 11.99 σ – much longer than in free speech (interview), where the mean was 6.96 σ (KENDALL 2013 note 5: 219). However, in Gut's corpus study, the mean length of run for L1 English read speech was 9.44 σ and in retelling 11.0 σ (GUT 2009: 99). These differences might be due to the methodological differences in the studies, specifically, to the reading materials used. This interpretation is supported by the fact that the speech rate measures in the L1-E corpus (cf. Table 3a) were more moderately different from Kendall's, where the mean articulation rate was 4.44 σ/sec, and mean speaking rate was 3.71 σ/sec, and very close to Gut's, where L1 reading articulation rate was 5.6 σ/sec and the speech rate was 4.1 σ/sec.

Table 3a. *L1-E corpus*: Means and variability measures for the *speech-rate* variables.

	Mean	Std. Dev.	Variance	Skewness		Kurtosis	
				Statistic	Std. Error	Statistic	Std. Error
Speaking RATE	4.087	.207	.043	-.137	1.014	1.503	2.619
Speaking rate PRUNED	4.155	.201	.040	.951	1.014	1.845	2.619
Articulation rate	5.219	.242	.059	1.945	1.014	3.802	2.619
Phonation–time ratio (AT/ ST)	78.333	2.305	5.311	-.668	1.014	-1.201	2.619
PACE (stressed syllables / min)	111.032	8.681	75.353	1.182	1.014	2.274	2.619
SPACE (stressed / all words)	.555	.018	.000	1.923	1.014	3.735	2.619

Table 3b. *L1-E corpus*: Means and variability measures for the *pause* variables.

	Mean	Std. Dev.	Variance	Skewness		Kurtosis	
				Statistic	Std. Error	Statistic	Std. Error
Number of pauses	37.75	3.403	11.583	-1.199	1.014	1.979	2.619
Pauses – total duration	15.084	2.388	5.704	1.720	1.014	3.082	2.619
Pause RATIO (% of sp. time)	21.667	2.305	5.311	.668	1.014	-1.201	2.619
Pauses per 100 syllables	13.337	1.146	1.312	-1.718	1.014	3.028	2.619
Pauses per minute	32.616	1.656	2.743	-1.129	1.014	1.379	2.619
No of pauses within IU	.75	1.500	2.250	2.000	1.014	4.000	2.619
Duration of pauses within IU	.247	.494	.244	2.000	1.014	4.000	2.619

Table 3c. *L1-E corpus*: Means and variability measures for the *disfluency* variables.

	Mean	Std. Dev.	Vari-ance	Skewness		Kurtosis	
				Statistic	Std. Error	Statistic	Std. Error
Number of disfluencies	3.00	1.826	3.333	.000	1.014	-3.300	2.619
Disfluencies – total duration	1.153	.879	.773	.546	1.014	-.622	2.619
Disfluencies % of sp. time	1.633	1.168	1.365	.163	1.014	-1.854	2.619
Disfluencies per 100 syllables	1.056	.637	.406	.039	1.014	-3.005	2.619
Disfluencies per minute	2.602	1.634	2.668	.363	1.014	-1.652	2.619
Degree of hesitancy (DOH)	8.183	5.511	30.371	.556	1.014	-1.099	2.619

Table 4. *L1-E corpus*: The frequencies of the *disfluency types*.

Total IU in L1-E corpus = 218	Frequency	Valid %	Cumulative %
No disfluencies	**206**	94.5	94.5
[um] pause filler	**0**	.0	94.5
false start or repetition	**7**	3.2	97.7
[er] pause filler	1	.5	98.2
prolonged vowel	**1**	.5	98.6
prolonged consonant	**1**	.5	99.1
other vocalizations	**1**	.5	99.5
explicit comment	**1**	.5	100.0
∑	**=12**	(5.5%)	

3.2.1. EFL vs. L1-E corpus – Comparisons. Compared to EFL, L1 English had a somewhat shorter mean length of run, with a shorter duration. This finding is contrary to some previous research, as, for instance, Gut (2009) found a greater difference in read English, with L2 speakers producing a shorter mean length of run (7.3 σ) than the L1 speakers (9.44 σ).

Means comparisons (ANOVA) showed statistically significant differences between L1 English speakers and EFL speakers for several variables in all the three categories. With respect to speech rate, both unpruned and pruned speaking rate, as well as the articulation rate were significantly different (all p=.001), higher in L1 English. The other speech rate variables did not show statistical significance. Of the pause variables, the pause ratio to speaking time was lower in EFL than in L1-E (cf. Table 1b – Table 3b), and with a notably greater variability. EFL speakers produced fewer pauses per 100 syllables than L1-E speakers, again with notably greater variability, and also produced fewer pauses per minute than L1-E speakers. These differences, however, did not reach statistical significance, except for the number of pauses per minute, which approached significance (p=.066).

More importantly, EFL speakers produced many more IU-internal pauses, since the mean number of such pauses was five times higher in the EFL corpus than in the L1-E corpus (cf. Table 1b vs. Table 3b). This difference only approached statistical significance (p=.077), which could be due to very high variability between

EFL speakers. The total number of within-IU pauses was proportionally very different, as they occurred in 5.6% (N=73) IUs in the EFL corpus, but only 1.4% (N=3) IUs in the L1-E corpus. These findings resonate with some previous research results, where L2 speakers were found to make pauses within IUs more frequently and for a longer time than L1 speakers (De Jong 2016; Tavakoli 2011).

Lastly, the number and duration of disfluencies came close to statistically significant differences (p=.061 and p=.073 respectively), and so almost did the DOH variable (p=.088). This lack of clear statistical significance, despite obvious differences in the numbers and mean values (cf. Table 1c – Table 3c), could also be ascribed to the high variability in the EFL data.

3.3. L1-S corpus – Temporal properties. The statistics of the L1 Serbian reference corpus are summed up in *Table 5*, with variables organized into three thematic categories.

The L1 Serbian reference corpus, like the L1 English one, was smaller than the primary EFL corpus, and yet the disfluency variables had conspicuously high values (Table 5c). The number of disfluencies identified was surprisingly high (14%), actually proportionally higher than in the EFL data, and also rather varied, as shown in *Table 6*.

Table 5a. *L1-S corpus*: Means and variability measures for the *speech-rate* variables.

	Mean	Std. Dev.	Vari-ance	Skewness		Kurtosis	
				Statistic	Std. Error	Statistic	Std. Error
Speaking RATE	5.905	.385	.148	1.694	1.014	2.796	2.619
Speaking rate PRUNED	6.028	.428	.183	1.614	1.014	2.438	2.619
Articulation rate	7.168	.504	.254	.954	1.014	1.949	2.619
Phonation–time ratio (AT/ ST)	82.425	2.215	4.906	.553	1.014	1.643	2.619
PACE (stressed syllables / min)	116.311	7.211	52.004	1.095	1.014	-.029	2.619
SPACE (stressed / all words)	.6139	.003	.000	-.008	1.014	.091	2.619

Table 5b. *L1-S corpus*: Means and variability measures for the *pause* variables.

	Mean	Std. Dev.	Vari-ance	Skewness		Kurtosis	
				Statistic	Std. Error	Statistic	Std. Error
Number of pauses	42.25	2.630	6.917	1.443	1.014	2.235	2.619
Pauses – total duration	14.460	2.143	4.591	1.126	1.014	.856	2.619
Pause RATIO (% of sp. time)	17.576	2.215	4.906	-.553	1.014	1.643	2.619
Pauses per 100 syllables	8.673	.415	.172	.649	1.014	-1.130	2.619
Pauses per minute	30.680	1.904	3.625	-.771	1.014	1.657	2.619
No of pauses within IU	5.25	4.573	20.917	-.196	1.014	-3.202	2.619
Duration of pauses within IU	1.337	1.276	1.628	.539	1.014	-.617	2.619

Table 5c. *L1-S corpus*: Means and variability measures for the *disfluency* variables.

	Mean	Std. Dev.	Vari-ance	Skewness		Kurtosis	
				Statistic	Std. Error	Statistic	Std. Error
Number of disfluencies	7.25	2.217	4.917	.482	1.014	-1.700	2.619
Disfluencies – total duration	1.629	.555	.308	-1.768	1.014	3.318	2.619
Disfluencies % of sp. time	2.006	.755	.569	-1.206	1.014	1.963	2.619
Disfluencies per 100 syllables	1.500	.472	.222	.534	1.014	-1.390	2.619
Disfluencies per minute	5.350	1.815	3.293	.083	1.014	-3.912	2.619
Degree of hesitancy (DOH)	17.358	5.503	30.284	.086	1.014	-1.476	2.619

Table 6. *L1-S corpus*: The frequencies of the *disfluency types.*

Total IU in L1-S corpus = 229	Frequency	Valid %	Cumulative %
No disfluencies	197	86.0	86.0
[um] pause filler	0	.0	86.0
explicit comment	**0**	.0	86.0
false start or repetition	**10**	4.4	90.4
prolonged vowel	**9**	3.9	94.3
prolonged consonant	**9**	3.9	98.3
[er] pause filler	2	.9	99.1
other vocalizations	**2**	.9	100.0
∑	**=32**	(14%)	

3.3.1. L1-S CORPUS VS. EFL AND L1-E – COMPARISONS. The mean length of run in L1 Serbian was 9.146 σ/ IU, almost twice the length in either L1 English or the EFL corpus, with a notable variability between speakers (SD=.323, Variance=.104). However, the mean duration of the run was 1.556 sec (SD =.148; Variance = .022), almost identical to the mean duration of the run in the EFL corpus, and only moderately longer than in L1 English. The between-speaker variability for the mean duration of the run was also notably lower than for other measurements, indicating that this property – a great number of syllables per run without a significant increase in duration time – may be a property of the speakers' L1 variety of Serbian. This is further supported by the finding that the average syllable duration was the shortest in L1 Serbian, unpruned (=.170sec) and pruned (=.137sec), than in EFL (unpruned=.277sec; pruned=.214sec) and in L1 English (unpruned=.245sec; pruned=.188sec).

All the speech rate variables showed highly statistically significant differences (ANOVA) between the EFL data and the L1 Serbian data – the speaking rate, pruned and unpruned, and the articulation rate (all p=.000), all much higher in L1 Serbian than in EFL (cf. Table 5 vs. Table 1), especially the pruned speaking rate and the articulation rate.

While it was expected, based on previous research with L2 learners, that EFL speakers would have a slower speech rate and articulation rate (GUT 2009;

De Jong – Mora 2019) than L1 English speakers, it was not expected that such a remarkable difference would be found between two groups of L1 speakers, English and Serbian (cf. Table 5 vs. Table 3). Such a high speech rate, unpruned and pruned, as well as a very high articulation rate in the reading style of the L1 Serbian speakers could be related to their particular mother-tongue variety, a Southern Serbian dialect (Prizren – Južna Morava, the urban Niš variety). This indeed resonates with folk linguistic perceptions of Southern Serbian speech being fast-paced, which the findings of this analysis seem to corroborate. Probably for the same reason, the PACE variable showed a significant difference (p=.000) between L1 Serbian and EFL, as well, i.e., a higher number of stressed syllables per minute, although no difference was found for the SPACE variable, i.e., the number of stressed words divided by the total number of words. Contributing to this interpretation, regarding the pause variables, a statistically significant difference was found for the number of silent pauses per 100 syllables (p=.001), notably higher in EFL than in L1 Serbian.

The means comparisons (ANOVA) of the L1 English and L1 Serbian data support this interpretation, too. A statistically significant difference was found for the pause/speaking time ratio (p=.047), higher in English than in Serbian, as well as for the number of pauses per 100 syllables (p=.000), much higher in English than in Serbian. No statistically significant difference was found for the number of pauses per minute, where both the means and the variance measures were much closer. However, statistically significant differences were found for all the speech rate variables (all p=.000), with much higher values in Serbian. The mean length of run was also higher in Serbian, both in syllables per IU (p=.000) and in seconds (p=.028), and so was the phonation-time ratio (p=.047). The SPACE variable also showed a significant difference (p=.001).

Further supporting the interpretation that our L1 Serbian corpus was characterized by peculiar temporal properties, the univariate analysis of variance for all the three corpora merged showed that for the number of pauses per minute the main influence of the L1-or-L2 speaker variable was strong, with no influence of the variable English or Serbian language. In other words, more pauses per minute were produced by both English and Serbian L1 speakers, while EFL speakers produced fewer pauses per minute than either group of native speakers. However, for the number of pauses per 100 syllables, the main influence was exerted by the variable of language, with no influence of the variable L1 or L2 speaker. This indicates that fewer pauses per 100 syllables was characteristic of the Serbian language variety analysed in this study.

However, the most striking finding in the L1 Serbian corpus were the disfluency variables, proportionally more numerous than even in the EFL corpus, and much more numerous than in L1 English. Contrary to our expectations, no disfluency variables showed a statistically significant difference between EFL and L1 Serbian (ANOVA), as the DOH was almost identical in L1 Serbian as in the EFL, both more than twice as high as the DOH in L1 English. The number of disfluencies per minute was also very similar in EFL and L1 Serbian (cf. Table 5c vs. Table 1c), while the number of disfluencies per 100 syllables was lower in L1 Serbian, but with no statistical significance. On the other hand, the means com-

parisons (ANOVA) between the L1-E and L1-S data showed statistically significant differences for the degree of hesitancy (p=.057), and for the total number of disfluencies (p=.025) and the number of disfluencies per minute (p=.065). Such high measures for the disfluency variables in L1 Serbian could be related to the high speech rate, but the results are inconclusive in this respect and require further research, particularly considering that the total number of within-IU pauses in the L1-S corpus, 9.2% of all the IUs produced (N=21), was proportionally the highest compared to 5.6% in the EFL and only 1.4% in the L1-E corpus.

3.4. Listeners' perception of fluency. In the perception part of the study, the evaluations of the two rater groups – untrained listeners and trained EFL teachers – were treated as separate variables, in order to observe the possible differences between their constructs of fluency, and the possibly different temporal aspects of speech their rating may be related to.

Indeed, some differences were observed. While high correlations were found between the students' and the teachers' ratings for all the three corpora merged (rho=.673*, p=.000), as well as for the EFL corpus (rho=.606**, p=.006), no correlation was found between their ratings of the L1 corpora, Serbian or English. *Table 7* shows the means and several variability measures for the two groups of raters for each corpus – the EFL, L1-E and L1-S – and for all the three merged together. For comparison, as there was no correlation between their ratings of L1-E and L1-S, the table also presents students' and teacher's ratings for all the L1 speakers together, irrespective of the language (English or Serbian).

Table 7. Means and variance measures of the fluency ratings**.**

	Mean	St. Dv.	Range	Min.	Max.	Variance	Skewness		Kurtosis	
								Std. Error		Std. Error
EFL corpus										
Students' rates	3.310	.568	1.97	2.21	4.18	.322	-.343	.524	-.826	1.014
Teachers' rates	3.724	.728	2.38	2.50	4.88	.531	-.098	.524	-1.191	1.014
L1-E corpus										
Students' rates	4.733	.096	.200	4.59	4.79	.009	-1.875	1.014	3.530	2.619
Teachers' rates	4.969	.063	.125	4.88	5.00	.004	-2.000	1.014	4.000	2.619
L1-S corpus										
Students' rates	3.558	.187	.440	3.38	3.82	.035	1.260	1.014	2.310	2.619
Teachers' rates	3.000	.791	1.750	2.25	4.00	.625	.632	1.014	-1.700	2.619
All 3 corpora merged										
Students' rates	3.557	.697	2.58	2.21	4.79	.485	.123	.448	-.350	.872
Teachers' rates	3.801	.868	2.75	2.25	5.00	.753	-.160	.448	-1.203	.872
Both L1 corpora										
Students' rates	4.145	.643	1.41	3.38	4.79	.413	-.090	.752	-2.496	1.481
Teachers' rates	3.984	1.173	2.75	2.25	5.00	1.377	-.593	.752	-1.642	1.481

Means comparisons showed that student raters seem to have had a positive bias towards native speakers in general, both L1 Serbian and L1 English, since their mean rates for both groups of native speakers were higher than for the EFL speakers, and also higher than the teachers' mean rating of all the L1 speakers. This is confirmed by the statistically significant difference between the students' ratings of all L1 vs. EFL speakers (ANOVA, F=11.304, p=.002), while no significant difference was found between the teachers' ratings of native vs. non-native speakers. On the other hand, teacher raters seem to have rated English speech, both L1 and EFL, more favourably than the L1-S corpus, as that was the only comparison in which the teachers' rate was lower than the students'. This is confirmed by the statistically significant albeit moderate difference (ANOVA, F=4.548, p=.042) between the teacher's ratings of EFL and L1-E vs. L1-S.

3.4.1. Temporal properties and fluency. The linear regression and correlation analyses established a clear relationship between certain temporal properties and the listeners' evaluations of the speakers' fluency, irrespective of the language or the L1 vs. L2 speaker difference. Regarding the speech rate variables, the speaking rate correlated with the students' ratings (rho=.439*, p=.022), and so did the articulation rate (rho=.433*, p=.024), but not with the teachers' ratings. The positive correlation coefficients indicate that student raters appreciated faster-paced speech as a sign of fluency more than teacher raters.

On the other hand, teacher raters (but not student raters) considered pauses at the end of IUs an indicator of higher fluency, judging by the positive correlation coefficient (rho=.052*, p=.030). This finding resonates with Marković and colleagues (2015), who observed that adding pauses between intonational phrases in synthesized speech contributed to better performance, i.e., listeners' evaluations.

However, both groups of raters considered within-IU pauses, silent or filled, as a clear sign of disfluency. The number of pauses within IUs correlated both with the teachers' (rho=-.446*, p=.020) and the students' (rho=-.504**, p=.007) ratings, and so did the duration of such within-IU pauses (teachers: rho=-.402*, p=.038; students: rho= -.498**, p=.008). The negative coefficients indicate that fewer pauses within IUs, and shorter pauses were interpreted as an indicator of higher fluency. Similarly, all the variables related to disfluencies showed strong correlations with both students' and teachers' ratings, as strong predictors of disfluency. *Table 8* shows the correlation matrix for this group of variables and the fluency ratings.

The negative correlation coefficients indicate that all raters interpreted disfluency elements, their number, duration, and frequency, as indicators of a lower degree of fluency, in line with previous research (e.g., Comeaux – Thomson 2019; Kahng 2018; Shea – Leonard 2019)). Some disfluency types particularly correlated with fluency ratings. Both groups of raters considered false starts and repetitions as indicators of disfluency (teachers: rho=-.508**, p=.007; students: rho=-.474*, p=.013), as well as consonant prolongations (teachers: rho=-.435*, p=.023; students: rho=-.541**, p=.004) The students – but not the teachers – found the [er] filler as a sign of disfluency (rho=-.498**, p=.008). The teachers also interpreted vowel drawls and vocalizations as a disfluency sign (rho=-.447*, p=.020), with the correlation approaching statistical significance with student raters, too (rho=-368, p=0.59).

Table 8. Correlation matrix (Spearman's rho) for the disfluency variables and fluency ratings.

		Teacher	Student	Disfluencies					
		rates	rates	Number	duration	% of Sp, time	per 100 syll.	per minute	DOH
Teacher rates	Coefficient	1.000	.673**	-.662**	-.502**	-.444*	-.529**	-.601**	-.580**
	Sig.	.	.000	.000	.008	.020	.005	.001	.002
Student rates	Coefficient	.673**	1.000	-.684**	-.528**	-.569**	-.739**	-.709**	-.709**
	Sig.	.000	.	.000	.005	.002	.000	.000	.000
Number	Coefficient	-.662**	-.684**	1.000	.856**	.853**	.930**	.934**	.940**
	Sig.	.000	.000	.	.000	.000	.000	.000	.000
Duration	Coefficient	-.502**	-.528**	.856**	1.000	.947**	.779**	.690**	.781**
	Sig.	.008	.005	.000	.	.000	.000	.000	.000
% of Sp, time	Coefficient	-.444*	-.569**	.853**	.947**	1.000	.848**	.781**	.863**
	Sig.	.020	.002	.000	.000	.	.000	.000	.000
per 100 syllables	Coefficient	-.529**	-.739**	.930**	.779**	.848**	1.000	.945**	.935**
	Sig.	.005	.000	.000	.000	.000	.	.000	.000
per minute	Coefficient	-.601**	-.709**	.934**	.690**	.781**	.945**	1.000	.966**
	Sig.	.001	.000	.000	.000	.000	.000	.	.000
DOH	Coefficient	-.580**	-.709**	.940**	.781**	.863**	.935**	.966**	1.000
	Sig.	.002	.000	.000	.000	.000	.000	.000	.

**. Correlation is significant at the 0.01 level (2-tailed) *. Correlation is significant at the 0.05 level (2-tailed).

Observing only the EFL corpus, significant correlations (Spearman's) were observed between the teacher's rates and false starts and repetitions (rho= -.574*, p=.010), the [er] filler (rho= -.470*, p=.042), and other vocalizations (rho= -.571*, p=.011). For students' fluency ratings, correlations were found for consonants prolongations (rho= -.566*, p=.012), and for false starts and repetitions it approached statistical significance (rho= -.429*, p=.067). The negative correlation coefficients suggest that higher fluency ratings were related to fewer disfluencies. However, fewer and weaker correlations in the students' ratings indicate that teacher raters considered particular disfluencies in EFL speech as stronger indicators of disfluency than student raters did.

The regression analyses supported these findings further. The predictors of pauses within IUs and the duration of such pauses both showed a very high significance (p=.006, p=.009 respectively), with F= 8.010 (p=.012) in the linear regression analysis, and F= 7.469 (p=.014, ANOVA) in the logarithmic analysis in students' ratings, while the duration of the pauses within IU was also significant in teachers' ratings (linear F=6.890, p=.018; logarithmic F=6.389, p=.022). *Figure 1* shows the scatterplots for the variables pause within the IU (left) and the duration of the pause within IU (right), illustrating the observation that fewer within-IU pauses related to higher fluency ratings, and, to a lesser degree, so did their shorter duration.

In the regression analysis, the disfluency variables showed the highest significance, too. For students' ratings, all the predictors showed significance: the number of disfluencies (p=.003), their duration (p=.057), the ratio of disfluencies

as the % of the total speaking time (p=.006), and primarily the number of disfluencies per 100 syllables and per minute (both p=.000), as well as the DOH (p=.000), with the F=8.769 (p=.001, ANOVA). All these predictors showed a very high significance in teachers' ratings as well (all p=.000, the duration of disfluencies p=.001), with the F=5.392 (p=.007, ANOVA). *Figure 2* shows the estimated curve (linear, logarithmic, logistic regression) in the scatterplot of the variable Degree of hesitancy (DOH) for teachers' ratings (left) and students' ratings (right), illustrating the observation that lower DOH was related to higher fluency ratings by both groups of raters.

Figure 1. Regression analysis (curve estimation) – scatterplots of the variables pause within the IU (left) and the duration of the pause within IU (right).

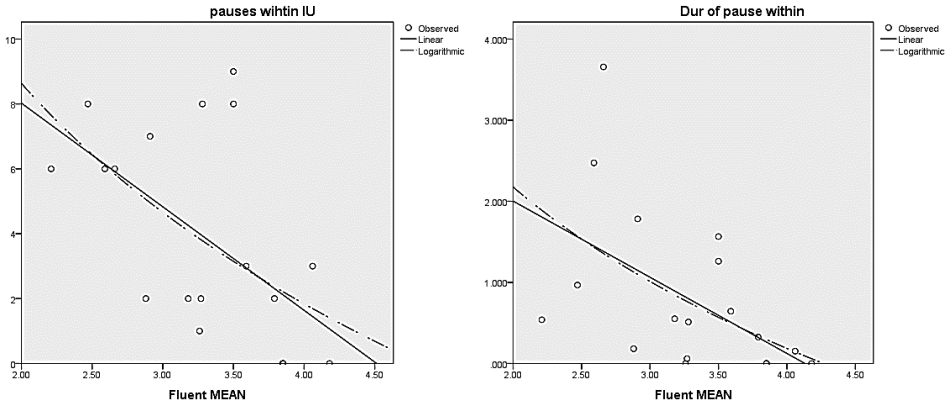
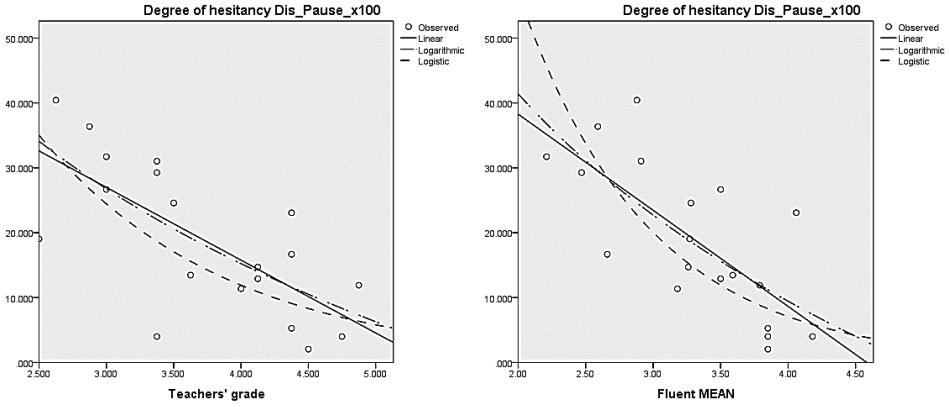


Figure 2. Regression analysis (curve estimation - linear, logarithmic, logistic) – the scatterplot of the variable Degree of hesitancy (DOH) for teachers' (left) and students' ratings (right).



4. Conclusion. The results of the quantitative corpus analysis showed that L2 English read speech differed from both L1 English and L1 Serbian. It was characterized by lower speaking rate and articulation rate values compared to L1 English, and, particularly, to L1 Serbian. The EFL speakers produced fewer pauses

than L1 English speakers, a somewhat longer mean run, fewer pauses overall, but significantly more pauses within IUs, and notably more disfluencies. EFL reading differed from L1 Serbian, too, as the latter was characterized by a very long mean length of run without a longer duration, by much higher speech rate and articulation rate, and fewer pauses per 100 syllables. Therefore, it can be said that the EFL reading did not exhibit a significant L1 influence with respect to speech rate variables, but did for some pause variables, e.g., fewer pauses compared to L1-E. Also, the somewhat longer mean length of run in EFL than in L1-E, which is unusual considering previous research findings, can be interpreted as a sign of L1 influence.

Some of the phonetic properties found in the L1-S corpus, such as very high speech rate and articulation rate, and long mean runs, as well as the shortest average syllable duration, could be attributed to the participants' specific variety of (Southern) Serbian, the urban Niš variety of the Prizren – Južna Morava dialect. On the other hand, the smaller number of pauses between IUs and numerous within-IU pauses and disfluencies cannot be explained by the specific dialectal temporal properties.

Some temporal properties were found to be significant predictors of (dis)fluency, irrespective of the language, and of the native vs. non-native speaker difference. Both student and teacher raters considered within-IU pauses and their duration as a clear sign of disfluency, as well as all the disfluency variables: their number, duration, ratio to the speaking time, and the degree of hesitancy (DOH) all correlated with lower fluency rates in both groups of raters. Specifically, false starts and repetitions were considered as indicators of disfluency by both groups of raters, although in several previous studies the authors pointed out that self-repairs of this kind need not necessarily be signs of disfluency (Götz 2013; Lennon 1990; Segalowitz 2010). This can be explained by the speaking modality investigated in this study – read speech, where, in difference to free or conversational speech, false starts and repetitions were not regarded as a useful self-repair technique, i.e., a "communicative signal", but, rather, as a "symptom" of disfluency (Crible 2018: 3).

An important finding of this study is also the difference observed between student raters, who considered faster-paced speech as a sign of fluency, and teacher raters, who considered pauses at the end of IUs an indicator of fluency. This is reflected in the fact that student listeners rated native speakers, both L1 Serbian and L1 English, overall more favourably than the EFL speakers, whose speech-rate variables were the lowest. On the other hand, teachers rated L1 Serbian speech less favourably, as it was characterized by long mean runs, fewer pauses, and a faster speaking rate. These findings support the claim that fluency is "in the ears of the listener", i.e., a listener construct, as the two groups of raters obviously had different expectations with respect to fluency.

Finally, although this study did not investigate fluency as related to EFL learners' proficiency levels, our findings do pose two questions related to pedagogical implications and language teaching. De Jong and colleagues (2012b) point out that if fluency is understood as a listener construct, it raises the question of L2 oral proficiency assessment. Our findings about the listener's perception of fluency also suggest that in language teaching and testing, the content of the notion of fluency should be explicitly defined and specified, in a way transparent to both students and teachers (and teacher-trainers).

Secondly, Gut (2009: 109–110) states that non-native speakers' fluency varies with speaking style, and is much higher in reading than in free speech, due to a lesser cognitive load (planning ahead for the content and the form); therefore, temporal features of speech can be taken as indicators of cognitive fluency and the proceduralization of knowledge; i.e., "utterance fluency" is indicative of "cognitive fluency" (Segalowitz 2010; 2016).

However our findings point to a somewhat different conclusion. Similarly to previous research, they indeed showed that EFL students' reading was more disfluent than L1 English speakers', with several disfluency characteristics identified. However, this does not explain our findings for the referent L1 Serbian corpus of read speech. As pointed out above, the smaller number of pauses between IUs and numerous within-IU pauses and disfluencies in L1 Serbian cannot be explained either by the specific dialectal temporal properties of these students' mother tongue, nor by the lack in the cognitive fluency or proficiency in their mother tongue. This raises the question of whether the disfluency characteristics of the EFL corpus should be attributed to the students' low proficiency in English, or their lower cognitive fluency, which would require comparisons across various speaking modalities. Or, apparently, it could be explained by the students' lack of a very specific oral sub-skill – that of reading aloud.

Our findings suggest that reading aloud, which has long been considered a traditional, conservative, and therefore undesirable type of classroom activity, seems to be regaining relevance today, as a skill that cannot be taken for granted even in speakers with a rather high level of education and otherwise high language proficiency. Moreover, in the Serbian formal educational contexts there is increasing anecdotal classroom evidence that speakers' oral performance – not only in L2 but also in L1 – has changed. While some decades ago students were, generally, much more fluent in reading than in spontaneous speech – as shown in the previous fluency research presented above – today the reverse seems to be gaining ground. Both L1 and L2 teachers report that some of their students can speak "smoothly and effortlessly" enough, but cannot really read fluently. In this light, reading aloud seems to be gaining renewed relevance as a literacy skill, in both L1 Serbian and EFL, requiring more attention from both L1 and L2 teachers and researchers.

REFERENCES

Boersma Paul, David Weenink. *Praat: Doing Phonetics by Computer*. 2013. (Version 5.2.04, 1992–2010) [Computer software]. (11.11.2017) <http://www.praat.org>

Clark, Herbert, Jean Fox Tree. Using *uh* and *um* in Spontaneous Speaking. *Cognition* 84 (2002): 73–111.

Comeaux, Katie C., Ron I. Thomson. Towards a Deeper, *uh*, Understanding of, *um*, L2 Fluency and its [750 ms] Silence. John Levis, Charles Nagle, Erin Todey (Eds.). *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conferenc*e, Ames, IA, September 2018. Ames, IA: Iowa State University, 2019, 106–115.

Crible, Ludivine. *Discourse Markers and (Dis)fluency: Forms and functions across languages and registers*. Amsterdam – Philadelphia: John Benjamins Publishing Company, 2018.

Crible, Ludivine, Amandine Dumont, Iulia Grosman, Ingrid Notarrigo. *Annotation Manual of Fluency and Disfluency Markers in Multilingual, Multimodal, Native and Learner Corpora. Version 2.0*. Technical report, Université catholique de Louvain and Université de Namur, 2016.

De Jong, Nivja. Predicting Pauses in L1 and L2 Speech: The Effects of Utterance Boundaries and Word Frequency. IRAL – *International Review of Applied Linguistics in Language Teaching* 54/2 (2016): 113–132. DOI 10.1515/iral-2016-9993

De Jong, Nivja, Hans Rutger Bosker. Choosing a Threshold for Silent Pauses to Measure Second Language Fluency. *Proceedings of Disfluency in Spontaneous Speech, DiSS 2013, KTH Royal Institute of Technology* 54/1 (2013): 17-20.

De Jong, Nivja, Joan C. Mora. Does Having Good Articulatory Skills Lead to More Fluent Speech in First and Second Languages? *Studies in Second Language Acquisition* 41 (2019): 227–239. doi:10.1017/S0272263117000389

De Jong, Nivja H., Margarita P. Steinel, Arjen Florijn, Rob Schoonen, Jan H. Hulstijn. Facets of Speaking Proficiency. *Studies in Second Language Acquisition*, 34/1 (2012a): 5–34. https://doi.org/10.1017/S0272263111000489

De Jong, Nivja H., Margarita P. Steinel, Arjen Florijn, Rob Schoonen, Jan H. Hulstijn. Linguistic Skills and Speaking Fluency in a Second Language. *Applied Psycholinguistics* 34/5 (2012b): 893–916.

Derwing, Tracey, Marian Rossiter, Murray Munro, Ron Thomson. Second Language Fluency: Judgments on Different Tasks. *Language Learning* 54/4 (2004): 655–679.

Eren, Ömer, Mehmet KILIÇ, Erdoğan Bada. Fluency in L2: Read and Spontaneous Speech Pausing Patterns of Turkish, Swahili, Hausa and Arabic Speakers of English. *Journal of Psycholinguistic Research* 51/2 (2022): 237–253.

Fillmore, Charles. On Fluency. Charles J Fillmore, Daniel Kempler, William S-Y. Wang (eds.) *Individual Differences in Language Ability and Language Behaviour.* New York: Academic Press, 1979, 85–101.

Goldman-Eisler, Frieda. A Comparative Study of Two Hesitation Phenomena. *Language and Speech* 4/1 (1961a): 18–26.

Goldman-Eisler, Frieda. The Distribution of Pause Durations in Speech. *Language and Speech* 4/4 (1961b): 232–237.

Goldman-Eisler, Frieda. The Significance of Changes in the Rate of Articulation. *Language and Speech* 4/3 (1961c): 171–174.

Goldman-Eisler, Frieda. *Psycholinguistics Experiments in Spontaneous Speech*. London: Academic Press, 1968.

Götz, Sandra. *Fluency in Native and Non-native English Speech.* Amsterdam – Philadelphia: John Benjamins Publishing Company, 2013.

Götz, Sandra. Filled Pauses Across Proficiency Levels, L1s and Learning Context Variables: A Multivariate Exploration of the Trinity Lancaster Corpus Sample. *International Journal of Learner Corpus Research (IJLCR)* 5/2 (2019): 159–180.

Gut, Ulrike. *Non-Native Speech: A Corpus-Based Analysis of Phonological and Phonetic Properties of L2 English and German.* Frankfurt: Peter Lang, 2009.

Kahng, Jimin. The Effect of Pause Location on Perceived Fluency. *Applied Psycholinguistics* 39/3 (2018): 569–591.

Kendall, Tyler. *Speech Rate, Pause, and Sociolinguistic Variation: Studies in Corpus Sociophonetics.* New York: Palgrave Macmillan, 2013.

Koreman, Jacques. Perceived Speech Rate: The Effects of Articulation Rate and Speaking Style in Spontaneous Speech. *The Journal of the Acoustical Society of America* 119/1 (2006): 582–596.

Kormos, Judit J. *Speech Production and Second Language Acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates, 2006.

Kosmala, Loulou, Ludivine Crible. The Dual Status of Filled Pauses: Evidence from Genre, Proficiency and Co-occurrence. *Language and Speech* 65/1 (2022): 216–239.

Krivokapić, Jelena. Prosodic Planning: Effects of Phrasal Length and Complexity on Pause Duration. *Journal of Phonetics* 35/2 (2007): 162–179.

Krivokapić, Jelena. Prosodic Planning in Speech Production. Susanne Fuchs, Melanie Weihrich, Daniel Pape, Pascal Perrier (eds.). *Speech Planning and Dynamics.* München, Germany: Peter Lang, 2012, 157–190.

Lehiste, Ilse. Traces of Dialectal Background in Spoken Literary Serbian. *Južnoslovenski filolog* 56/1–2 (2000): 589–594.

Lennon, Paul. Investigating Fluency in EFL: A Quantitative Approach. *Language Learning* 40/3 (1990): 387–417.

Marković Maja, Bojana Jakovljević, Tanja Milićev, Nataša Milićević. The Role of Prosody in the Perception of Synthesized and Natural Speech. Andrey Ronzhin, Rodmonga Potapova, Nikos

Fakotakis (eds.). *Speech and Computer, SPECOM 2015, Lecture Notes in Computer Science,* vol. 9319. Cham: Springer, 2015, 446–453. <https://doi.org/10.1007/978-3-319-23132-7_55>.
Redford, Melissa. A Comparative Analysis of Pausing in Child and Adult Storytelling. *Applied Psycholinguistics* 34/3 (2013): 569–589.
Redford, Melissa. The Acquisition of Temporal Patterns. Melissa Redford (ed.). *The Handbook of Speech Production*. Boston: Wiley-Blackwell, 2015, 379–403.
Rossiter, Marian J. Perceptions of L2 Fluency by Native and Non-native Speakers of English. *Canadian Modern Language Review* 65/3 (2009): 395–412.
Segalowitz, Norman. *Cognitive Bases of Second Language Fluency.* New York: Routledge, 2010.
Segalowitz, Norman. Second Language Fluency and its Underlying Cognitive and Social Determinants. IRAL – *International Review of Applied Linguistics in Language Teaching* 54/2 (2016): 79–95.
Shea, Christine, Karen Leonard. Evaluating Measures of Pausing for Second Language Fluency Research. *Canadian Modern Language Review* 75 /3 (2019): 216–235.
Tavakoli, Parvaneh. Pausing Patterns: Differences between L2 Learners and Native Speakers. *ELT Journal* 65/1 (2011): 71–79.
Tomić, Kristina. Temporal Parameters of Spontaneous Speech in Forensic Speaker Identification in Case of Language Mismatch: Serbian as L1 and English as L2. *Comparative Legilinguistics* 32 (2017): 117–143.
Trouvain, Jürgen, Jacques Koreman, Attilio Erriquez, Bettina Braun. Articulation Rate Measures and Their Relation to Phone Classification in Spontaneous and Read German Speech. Jena-Claude Junqua, et al. (Eds.). *Proceedings of the Workshop Adaptation Methods for Speech Recognition: Sophia-Antipolis, France, August 29 - 30, 2001 [Isca ITR-workshop].* Sophia–Antipolis, 2001, 155–158.
Wagner, Petra, Jürgen Trouvain, Frank Zimmerer. In Defense of Stylistic Diversity in Speech Research. *Journal of Phonetics* 48 (2015): 1–12.
Word Count, online application, Font Keyboard LLC © 2022. <https://wordcounttools.com/>.

Татјана Пауновић

ТЕМПОРАЛНА СВОЈСТВА И ПЕРЦЕПЦИЈА ТЕЧНОСТИ ГОВОРА У ЧИТАЊУ У СРПСКОМ, ЕНГЛЕСКОМ И ЕНГЛЕСКОМ КАО СТРАНОМ ЈЕЗИКУ

Р е з и м е

Појам течности говора сматра се кључним у учењу језика. Ипак, упркос бројним истраживањима, још увек није прецизно утврђено шта слушалац перцепира као течност, било у матерњем или страном језику, а у српском језику, као и у учењу енглеског као страног језика у српском образовном контексту, ово питање до сада није истраживано.

У овом раду представљено је истраживање темпоралних својстава говора у корпусу текстова које су читали студенти енглеског као страног језика (EFL корпус), и мањим референтним корпусима које су читали изворни говорници енглеског (L1-E) и српског језика (L1-S). Квантитативна анализа корпуса обухватала је акустичка мерења основних темпоралних својстава, срачунавање сложених параметара, као и статистичку анализу више варијабли груписаних у три категорије: брзина говорења, паузе и „поштапалице", односно, вокални елементи без значења којима говорник испуњава паузе (енг. *disfluencies*). Две групе оцењивача слушале су снимљене узорке говора и оцењивале течност на петостепеној ликертовој скали – група студената без претходног образовања из фонетике или лингвистичко-методичких дисциплина, и три наставника енглеског као страног језика. Оцењивачима нису дате посебне инструкције или објашњења, са циљем да се утврди да ли ће се слушаоци ослањати на иста темпорална својства при оцењивању, и да ли ће се њихови концепти течности говора разликовати.

Резултати истраживања су показали, сасвим у складу са претходним истраживањима са ученицима енглеског језика различитог порекла и нивоа знања, да се EFL корпус разликовао од L1-E корпуса по нижим вредностима варијабли брзине говорења, као и варијабли у категорији

пауза, али уз бројније паузе унутар интонацијских целина, као и знајно бројније вокалне елементе који служе као поштапалице. Није утврђен утицај матерњег језика у домену варијабли брзине говорења, али јесте утврђен утицај у домену варијабли у категорији пауза и дужег просечног говорног низа између две паузе.

Резултати су показали и значајну повезаност између више темпоралних својстава и оцене течности говора, али и неке разлике између две групе оцењивача. Најзначајнији предиктори ниске оцене течности говора биле су паузе унутар интонацијских целина, код обе групе оцењивача, као и све поштапалице. Међутим, док су код наставника пузе на крају интонацијских целина биле значајан индикатор боље течности говора, код студената-оцењивача већа брзина говорења била је индикатор течнијег говора. Резултати овог истраживања су такође показали и нека темпорална својства у L1 S корпусу која се могу повезати са карактеристичним варијететом матерњег језика испитаника – урбаним нишким варијететом призренско-јужноморавског дијалекта – као што су висока брзина говорења, кратко трајање просечног слога, и веома дуги говорни низови између пауза, без повећања трајања низа.

Коначно, резултати су указали и на неке педагошке имипликације за наставу како страног тако и матерњег језика, на пример, да је приликом оцењивања говорне перформансе неопходно прецизно и експлицитно формулисати садржину појма флуентног, односно, течног говора, како за наставнике, тако и за ученике. Такође, резултати указују да је потребно посветити више пажње, како у настави тако и у истраживањима, вештини читања као засебној говорној вештини.

Универзитет у Нишу
Филозофски факултет
Департман за англистику
Ћирила и Методија 2, 18000 Ниш, Србија
tatjana.paunovic@filfak.ni.ac.rs

(Примљено: 1. септембра 2022;
прихваћено: 10. новембра 2022)