UNIVERZITA KARLOVA

Filozofická fakulta

BAKALÁŘSKÁ PRÁCE

Praha 2025 Julie Řihošková

Univerzita Karlova

Filozofická fakulta

Ústav anglického jazyka a didaktiky

Bakalářská práce

Julie Řihošková

Comparative analysis of disfluencies of native Czech speakers in their L1 and L2 in the English Teacher Corpus

Srovnávací analýza dysfluencí rodilých mluvčích češtiny v jejich L1 a L2 v korpusu učitelské angličtiny

Praha 2025

Vedoucí práce: PhDr. Tomáš Gráf, Ph.D.

Poděkování

Děkuji vedoucímu své práce, panu Gráfovi, za nejen jeho pomoc s touto prací, ale také především za to, že je skvělým vzorem pro desítky budoucích učitelů, kterým jde příkladem. Jeho entuziasmus, láska k učení a lidskost byly pro mě velkou inspirací nejen po dobu psaní práce pod jeho vedením, ale napříč mým studiem.

Prohlášení:

Prohlašuji, že jsem bakalářskou práci vypracoval/a samostatně, že jsem řádně citoval/a všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

I declare that the following BA thesis is my own work for which I used only the sources and literature mentioned, and that this thesis has not been used in the course of other university studies or in order to acquire the same or another type of diploma

V Praze dne …….. *podpis*

Souhlasím se zapůjčením bakalářské práce ke studijním účelům. I have no objections to the BA thesis being borrowed and used for study purposes.

Abstract:

This thesis examines the occurrence and nature of disfluencies in the spontaneous speech of Czech teachers of English, particularly false starts and repeats. Based on previous research on fluency, disfluency, and cross-linguistic influence, this study investigates whether speakers' patterns of disfluency in their first language (Czech) correspond to patterns in their second language (English), despite their high level of language proficiency. The theoretical part introduces key concepts of cognitive and temporal aspects of fluency, communicative functions of disfluency, and the role of language transfer in second language production. The analysis uses the English Teacher Corpus and compares two spontaneous monologues from each speaker, in their L1 and L2. Disfluencies were manually identified and categorised according to consistent methodological criteria. The results show that individuals who produce a higher number of false starts and repeats in Czech also produce a higher number of false starts and repeats in English. Additionally, the types of repeated words and the average length of disfluencies in both languages show notable similarities. The findings suggest that speech fluency habits developed in the first language persist even at advanced levels of second language proficiency. The study contributes to a more nuanced understanding of the speech fluency of L2 teachers and emphasises the complex interaction between language competence and cognitive speech planning.

Keywords: fluency, disfluency, false starts, repeats, teacher language, English Teacher Corpus, frequency comparison, English of Czech teachers of English, language transfer

Abstrakt:

Tato práce zkoumá výskyt a povahu dysfluencí ve spontánní řeči českých učitelů angličtiny, zejména plané začátky a opakování slov. Na základě předchozích výzkumů plynulosti, dysfluence a mezijazykového transferu tato studie zkoumá, zda dysfluence mluvčích v jejich mateřském jazyce (češtině) odpovídají těm v jejich druhém jazyce (angličtině), a to i přes jejich vysokou úroveň jazykové znalosti. Teoretická část představuje klíčové pojmy kognitivních a časových aspektů plynulosti, komunikační funkce dysfluence a roli jazykového transferu v produkci druhého jazyka. Analýza využívá korpus English Teacher Corpus a porovnává dva spontánní monology každého mluvčího v jeho L1 a L2. Dysfluence byly ručně zaznamenány a členěny podle konzistentních metodologických kritérií. Výsledky ukazují, že mluvčí, kteří v češtině produkují vyšší počet planých začátků a opakování, je produkují ve vyšší míře i v angličtině. Kromě toho jsou typy opakovaných slov a průměrná délka dysfluencí v obou jazycích velmi podobné. Výsledky naznačují, že návyky plynulosti řeči získané v prvním jazyce přetrvávají i na pokročilé úrovni znalosti druhého jazyka. Studie tak přispívá k lepšímu pochopení plynulosti řeči učitelů L2 a zdůrazňuje složitý vztah mezi jazykovou kompetencí a kognitivním plánováním řeči.

Klíčová slova: plynulost, dysfluence, plané začátky, opakování slov, učitelský jazyk, korpus učitelské angličtiny, porovnávání frekvence, angličtina českých učitelů angličtiny, jazykový transfer

Table of contents:

List of abbreviations:

L1 – first language

L2 – second language

FS – false starts

R – repeat

EFL – English as a foreign language

ETC – English Teacher Corpus

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

phw – per hundred words

List of tables:

List of figures:

1 Introduction

There are many aspects of language that one must perfect before being able to claim to have mastered it. Although a strong knowledge of grammar and an extensive vocabulary would be a presumed requirement for such a statement, other important factors, including fluency, are often neglected. And, hopefully, as a future full-time English teacher, one of my biggest goals should be to achieve - or at least strive for - perfection in the language. I have often encountered the opinion that Czech teachers of English sound like a textbook or an encyclopaedia and lack fluency in the language. So, the invitation to use and contribute to the English Teacher Corpus intrigued me. Although I went into the project not knowing what my bachelor's thesis topic would be, after listening back to the recordings of the three teachers I had made, a question arose. One of the teachers kept repeating themselves often, both in their L1 and L2, which almost gave the impression of disfluency. In contrast, the other teacher's speech sounded smooth, and despite frequent repeats in English, their speech still sounded much more natural, rather than hurting the impression of good fluency. Is there a correlation between the number of repeats and false starts a person makes in their L1 and L2? And how do they differ across those two languages?

When people speak, they often hesitate, repeating words, stopping mid-sentence, or even restarting their thoughts entirely. This happens not only in a second language, but even when speaking one’s mother tongue. And since those interviews, I have thought about this topic a lot. These repeats and false starts are interesting, as they could reveal the inner processes of planning, thinking, and searching for words that occur beneath the surface of fluent speech. In particular, I became curious about how these disfluencies behave when speakers move between languages they know well, especially between their native language and a language they use professionally, such as English. Having tutored, taught adults, and worked as a substitute English teacher at a gymnázium, I have had the chance to observe language not only as a learner, but also from the other side of the classroom. It made me realise how much pressure teachers often feel to speak fluently and confidently, and how easily some disfluencies, such as repeats or false starts, can affect the self-impression of proficiency. I started to wonder whether these natural hesitations really signal a lack of fluency, or whether they are simply a normal part of real-time speech planning that persists across languages. While researching existing literature, I noticed that many studies focus either on fluency and disfluency in first language (L1) speech or in second language (L2) speech, but much less often on the transfer between the two. Some researchers have studied whether fluency skills transfer from L1 to L2 (e.g., Derwing et al., 2009; Peltonen, 2018), but the specific mechanisms behind disfluencies, especially false starts

and repeats, have not been fully mapped across languages. Most importantly, I realised that few studies have focused on highly proficient non-native speakers, such as EFL teachers, whose use of English is advanced, but whose L1 processing habits may still influence their speech. This gap in research led me to focus on the question of disfluency transfer: do speakers who produce more disfluencies in their native language also tend to produce more in their second language? And if so, do the characteristics of those disfluencies, such as which words are repeated, or how long hesitations last, reflect each other between L1 and L2?

To study these questions, I work with recordings from the English Teacher Corpus (ETC), a unique project that records interviews with Czech EFL teachers both in English and Czech. Using these data, I analyse two spontaneous monologues, one in each language, and compare the occurrences of false starts and repeats. My aim is not only to measure how often disfluencies occur, but also to find out whether the structure and length of hesitation show some cross-linguistic patterns. At the end of this study, I hope to have a better understanding of how second-language fluency is shaped not only by external teaching and learning, but also by internalised habits from the speaker's first language. For the purposes of this study, I have the following research questions:

1) Do teachers who produce more disfluencies in Czech (L1) also produce more disfluencies in English (L2)?
2) Which word classes appear most often in one-word repeats in L1 and L2?
3) Are disfluencies longer in English (L2) than in Czech (L1)?

Based on previous studies suggesting partial transfer effects in fluency behaviour (e.g., Gráf, 2017; Derwing et al., 2009), I believe that a relationship will exist between disfluency patterns in the two languages: that speakers who produce more disfluencies in Czech will tend to do so in English as well, and that hesitation patterns (both in word class and disfluency length) will show similarities across languages.

2 Theoretical background

2.1 Fluency

As this paper focuses on disfluencies, it is essential to first address what fluency is. Fluency is a rather difficult topic to define. To measure it, a set of measurable variables that create a model of what we call "fluency" is needed, as we are trying to assign concrete, assessable parameters to an abstract concept. If we were to try to find a definition, after a quick search, we would discover that it is not a concept which everyone agrees on. For example, the Cambridge Dictionary provides a simple definition: "the ability to speak or write a language easily, well, and quickly." The Merriam-Webster Dictionary provides an even shorter answer: fluency is "the quality or state of being fluent". And when clicking on the word *fluent*, it states: "capable of using a language easily and accurately". While both dictionaries are presumably satisfactory for the average speaker to understand the meaning of the word, both leave the question - what does it mean to speak "well, easily and accurately"? To find a sufficient answer, it is worth considering the word itself. *Fluency* comes from the verb *to flow*, meaning to move smoothly, gracefully, and is often associated with water or time. In Czech, the word *plynulost* and its verb form *plynout* also have similar connotations of a continuous and peaceful movement of a body of water or of time itself. We can also consider the Czech word *zběhlý*, from the verb běžet (to run), which also indicates movement. The German version would be *die Flüssigkeit*, this time the word *der Fluss* (a river) or *fließen* (to flow) being directly incorporated within the term itself. These are examples of just three of the many languages that view fluency in a similar light. The only way to detect or document motion is to associate it with time. This suggests that it is a smooth, continuous matter that occurs over time. This knowledge leads us to wonder what is happening within that time, for example, how much I can fit into that time and what I can fit in, for instance, apart from the content language itself. However, there are many more ways to view the factors contributing to the impression of fluency. It constitutes many aspects, which can include knowledge of vocabulary and grammar, flow, continuity, automaticity, or smoothness of speech (Koponen & Riggenbach, 2000). Other factors suggested are speech rate, length, frequency, non-lexical fillers, and others (Rossiter et al., 2010, p. 584).

2.1.1 Cognitive fluency

Nevertheless, this is not the only approach to the question of fluency. Some researchers do not look at it from the perspective of temporal variables but see fluency, for example, as cognitive, as in the mental process necessary in generating speech. This includes how quickly and effectively a speaker can retrieve lexical items and construct grammatical structures. Linguists interested in cognitive fluency often look at hesitations, false starts, and self-

corrections as indicators of the effort involved in producing speech. In her research, Wray concludes that formulaic language is necessary to 'sound fluent'. In her book, *Formulaic Language and Lexicon*, she summarises its contents as follows: "The book culminates in a new model of lexical storage, which accommodates the curiosities of non-native and aphasic speech. It proposes that parallel analytic and holistic processing strategies are able to reconcile, on the one hand, our capacity for understanding and producing novel constructions using grammatical knowledge and small lexical units, and, on the other, our use of prefabricated material which, although less flexible, also requires less processing. The result of these combined operations is a language that is fluent and idiomatic, yet crafted for its referential and communicative purpose" (Wray, 2002, p. 1). These pre-stored pieces of language facilitate cognitive processes involved in language production and comprehension. As formulaic sequences are stored and retrieved as whole 'chunks' from memory, the brain does not always have to construct sentences from individual words or syntactic rules. This reduces the cognitive load during language production, allowing for more mental resources to be available for other aspects of communication, such as the content of the message or the social nuances of interaction. Therefore, for some researchers, fluency is increasingly understood not simply in terms of syntactic proficiency, expansive vocabulary, and the mechanics of speech rate and continuity, as stated previously. Instead, a growing body of researchers emphasise the cognitive processes involved in retrieving linguistic data. For those, cognitive agility[1] in accessing and deploying linguistic knowledge is essential to what constitutes fluency.

Having considered both the temporal and cognitive aspects of fluency, it is apparent that interruptions or breaks in this flow - disfluencies - play a crucial role in how speech is produced and perceived. It is, therefore, useful to examine disfluency as such and not to consider it merely as the 'opposite' of fluency.

2.2 Disfluencies – the opposite of fluency?

Now that it has been established that fluency is a broad subject involving various factors, it is time to shift to the topic of disfluency itself. As described above, the very term *disfluency* indicates it to be the opposite of fluency. It is noteworthy, however, that that is not necessarily true, and not all researchers share this view. As Williams (2022) writes, this is a relatively newly researched domain that has already expanded beyond the perception of disfluencies as merely a part of speech disorders and has begun to be addressed in more detail in linguistics and

[1] 'Cognitive agility' could be seen as rapid and efficient access to stored linguistic information and the seamless execution of this information in real-time communication.

psycholinguistic studies. He proposes that "[a]s social interactants, speakers have to consider the effect on the listener and beware of interruptions, non-sequiturs, mistaken inferences, and other hazards, and intervene in the stream of speech to change course when necessary" (Williams, 2022, p. 2). Williams believes that disfluencies often indicate new or difficult-to-articulate information in speech, serving pragmatic functions rather than merely being errors or flaws in speech. He notes that disfluencies (e.g., filled pauses and repetitions) can help listeners by signalling upcoming complex information or changes in speech, which can assist in guiding listeners' expectations and facilitating comprehension. As Arnold et al. (2003) similarly conclude, disfluencies, such as hesitations ("uh", "um") or elongations ("theee"), commonly occur when speakers refer to objects or concepts that are new in the discourse rather than familiar or previously mentioned. There are multiple ways to perceive disfluencies, and to study them, it is necessary to 'choose' the correct variables and approach.

The study of disfluencies, such as false starts and repeats, can be thought of as belonging closely to the area of cognitive fluency. Belz et al. show how disfluencies, such as self-repetitions and false starts, occur in both L1 and L2 speakers of German, demonstrating how cognitive processes affect language fluency. As they state, "Krashen (1988) identifies the overuse of this increased monitoring[2] during second language speech production as one reason for disfluencies. For native English, Schneider (2014, 243) investigates two-word chunks and concludes that 'the more likely two words are to co-occur, the less likely speakers are to place a hesitation between them'. Assuming that learners of a second language know fewer prefabricated chunks, we would expect them to produce more silent and filled pauses than native speakers" (Belz et al., 2017, p. 8). As apparent from this passage, Belz et al. connect the cognitive 'burden' of language production to the occurrence of disfluencies. Their research also emphasises the challenges faced by L2 speakers of a language when accessing and using linguistic structures as fluently as native speakers. According to Belz et al., disfluencies are phenomena that “do not add propositional content to an utterance” yet are crucial in the speaker's cognitive processing of language (2017, p. 118). For instance, filled pauses and self-repairs, including false starts and repeats, indicate the speaker's cognitive strain. They manage speech production issues and help re-plan utterances, which is critical for maintaining fluency

[2] 'Monitoring' in language production refers to the cognitive process where speakers supervise and adjust their speech output to ensure linguistic accuracy. This can involve detecting and correcting potential errors during speech, a mechanism that is mainly present in second-language learners. Overusing this monitoring can lead to increased disfluencies, such as pauses and hesitations, as learners consciously apply grammatical rules and search for appropriate and 'correct' lexical items.

(Belz et al., 2017, p. 119). Fox Tree conducted another study suggesting disfluencies to be a part of cognitive fluency, and it provides a focused examination of how such disfluencies impact the cognitive processing required for understanding subsequent speech. The study shows how speech disfluencies, such as false starts and repetitions, require higher cognitive effort from listeners, directly supporting their ability to process and interpret subsequent linguistic information effectively. She writes, "Listeners require additional processing time to integrate these disrupted elements into their ongoing interpretation of the speech, reflecting a significant cognitive load" (Tree, 1995, p. 709). This increased cognitive effort fits well with the idea of cognitive fluency, which focuses on how people use their mental processing abilities and strategies to handle and understand language in real time. These findings emphasise that disfluencies reflect the mental 'burden' of speech production. In addition to revealing the cognitive load, these disfluencies may also serve strategic purposes - an idea that will be explored next.

For a thorough analysis of what lies behind the feeling of being "fluent" in a language, these factors play a significant role.[3] Disfluencies, such as filled pauses and repeats, are sometimes used strategically by speakers to manage the timing of language production, giving themselves time to plan their speech (Ferreira et al., 2004, p. 722). This strategic use is reflected in the fact that these pauses and repetitions can provide a time window that allows speakers to organise their thoughts and continue their speech more coherently. The presence of disfluencies also affects the listener's processing of language. For example, filled pauses might signal to the listener that the speaker is searching for the correct lexical item or syntactic structure, which influences how listeners interpret the ongoing speech. This can help the listener adjust their expectations and processing in real time to accommodate the speaker's disfluencies (Ferreira et al., 2004, p. 723-724). Ferreira et al. state: "Disfluencies, which typically occur when the speaker is uncertain how to continue, form part of the linguistic input a listener must interpret. These include filled and unfilled pauses that affect the timing of substitution operations in the parser's processing, thus playing a strategic role in the management of speech production"

[3] An example of this would be person 1, who speaks very quickly but outwardly sounds like their speech is still smooth and fluid. Such a person, however, would very likely have more disfluencies produced when counted, such as false starts. Their words could, colloquially speaking, ‘overtake their thoughts’, and the so-called disfluencies would serve at that moment, for example, to buy more time so that their speech and thoughts would meet again. Alternatively, a person 2 may be very nervous, in which case their speech production process would be affected by this state. And it would depend solely on the particular individual how they react to such a situation, whether in such cases, for example, their speech slows down and they stutter less or, on the other hand, even more, or whether that individual instead starts speaking in a similar way as described in the example with person 1.

(2004, p. 722). It is not the subject of this paper to deal with the thought processes that lead to seemingly fluent or disfluent speech. Nevertheless, the reason why it is important to consider, for example, cognitive fluency when researching anything related to speech fluency, although this is a factor that is difficult to measure, is that the conclusions that are drawn from the research will always be influenced by these factors. Even merely having a respondent feeling 'under the weather' on a given day may lead to results that do not correspond to their actual speech, for example. For this reason, it is important to bear in mind that the data may not perfectly reflect actual speech patterns.

So far, disfluencies have been discussed only generally and by using examples. According to Williams (2022), six main types of disfluencies can be measured and subsequently analysed, including silent pauses, filled pauses, prolongations, repetitions, self-corrections, and false starts. In my work, I focus on repeats and false starts, fully acknowledging that there are others. In the next section, false starts and repeats will be discussed.

2.3 False starts and repeats

2.3.1 False starts

2.3.1.1 False starts according to Williams

A false start is defined as the speaker abandoning the initial utterance to start again or continue in a different direction. Williams (2022) explains that a false start consists of three components: the "problem source," an "editing phase" where the speaker pauses or uses filler words, and a "revision" that corrects or reorients the utterance (p. 215). Unlike errors requiring correction, false starts are often preventive adjustments made before the listener perceives a problem. Williams also emphasises Levelt's (1983) concept of "covert corrections," where the speaker identifies a problem and corrects it without significantly disrupting fluency (p. 215). These changes are less about maintaining grammatical accuracy and more about ensuring natural discourse flow.

False starts are an essential part of spoken language and reflect the active involvement of the speaker in monitoring their speech. As Kormos (2006, cited in Williams, 2022:235) argues, false starts reveal the cognitive processes of the speaker as they navigate planning and formulation problems. The complexity of false starts, especially for advanced learners, often

implies the speaker's attempt to meet higher discourse demands[4] while keeping communicative coherence.

2.3.1.2 Why false starts matter

False starts allow speakers to adjust their communication in real time, which means they can improve their clarity and accuracy. Williams (2022) stresses that false starts not only help speakers organise their thoughts but also benefit the listener as they promote "discourse coherence" (p. 217). This coherence, first described by Levelt (1983), emphasises that false starts help incorporate new ideas into ongoing communication, making speech more structured and comprehensible to listeners. Additionally, false starts are significant for language acquisition. Williams (2022) notes that beginner and intermediate learners rely on false starts to self-monitor and experiment with language structures, often to revise utterances to reflect more accurate linguistic forms (p. 235). According to Verhoeven (1989, cited in Williams, 2022), false starts function as "productive devices" that allow learners to try and improve their language use (p. 235). This repetitive process helps not only language development but also helps learners internalise the 'norms' of the target language over time. False starts can also reveal how speakers adapt their language to contextual demands. In spontaneous conversations, false starts occur more frequently because they allow speakers to adjust their utterances based on listener feedback or surrounding context. As Williams (2022) comments, this flexibility is crucial for effective communication in real-world settings (p. 217).

2.3.1.3 Why false starts occur

False starts occur mainly due to cognitive and linguistic difficulties during speech production. Williams (2022) explains that false starts often occur when speakers face high cognitive load, such as recalling complex or low-frequency vocabulary, structuring ideas, or planning grammatical forms (p. 217). This phenomenon corresponds with Skehan's (2009) findings that planning demands significantly affect fluency, especially for less proficient speakers (as cited in Williams, 2022:215). In addition, false starts often occur during tasks requiring unstructured responses, where the speaker must simultaneously both conceptualise and formulate their ideas. Williams (2022) points out that these problems are more pronounced in second language learners, who often lack the automaticity of native speakers and must rely on working memory to construct utterances (p. 235). This reliance on limited cognitive

[4] Discourse demands would be the challenges of organising ideas into a clear and meaningful flow. An example would be presenting an argument during a debate, where speakers must align content with context and audience expectations (Skehan, 2009, as cited in Williams, 2022:215).

resources increases the likelihood of disfluencies such as false starts. Social and contextual factors also contribute to false starts. Williams (2022) cites Bortfeld et al. (2001), who found that variables such as topic complexity, speaker role, and audience presence influence the frequency and nature of false starts (p. 217). These factors suggest that false starts are not merely linguistic errors but adaptive strategies shaped by the environment and the speaker's communicative goals.

2.3.1.4 Proficiency level and false starts

The frequency and nature of false starts vary significantly across levels of proficiency. Williams (2022) explains that beginner and intermediate students produce more false starts due to their limited automaticity and working memory, which prevent them from planning and producing fluent speech (p. 235). In contrast, advanced learners produce fewer false starts, reflecting their better fluency and linguistic accuracy. However, Williams points out that advanced learners' false starts are often better formulated and strategically used to increase discourse coherence (p. 235). As mentioned previously in the chapter “Disfluencies – the opposite of fluency?”, this could occur, for example, when a proficient speaker with a faster speech rate pauses and therefore bides more time to collect their thoughts. However, in general, lower-level speakers produce more disfluencies.

This approach is consistent with the findings of Verhoeven (1989, cited in Williams, 2022), who described false starts as "productive devices" that allow learners to experiment with language use and improve their speech (p. 235). As learners increase their proficiency, they rely less on false starts for self-control and more on other fluency strategies, such as pre-emptive planning and lexicalised filler devices.[5] Williams (2022) notes that this shift reflects the learner's increasing automaticity and ability to produce accurate and coherent speech without frequent interruptions (p. 235).

2.3.2 Repeats

2.3.2.5 Definition and naming them

Academics differ in their approach to classifying this phenomenon. What they agree with is that a sequence of identical elements constitutes a repetition. However, the main debate is whether to count only repeated words and phrases or also include the repetition of phonemes. Kjellmer (2008) defines repeats broadly as the occurrence of an identical spoken element immediately or almost immediately repeated within the same turn. His definition includes single sounds (“d-d-dog”), parts of words (“re- really”), whole words (“I I I don't know”), and

[5] such as “uhm” or filler words as “like”, “you know” or “I mean”.

sequences of words ("It is it is a big issue"). This corresponds with Williams (2020): "Repetitions are the immediate and identical repeat of spoken material just uttered, e.g., w-w-will, i-is, a a de-dependent phrase."

On the other hand, Clark and Wasow (1998) define repeats primarily as a speech repair mechanism that occurs when a speaker commits to *a word or phrase* but then suspends speech before restarting.[6] A similar definition can be found in a study by Gráf (2017), where repeats are defined as segments of speech (whole words and clusters of words) that are involuntarily repeated in close proximity without adding any propositional content to the message. This paper will classify and categorise repeats according to this second definition, i.e., it will consider only repetitions of whole words and word clusters as repeats. The repeat of individual phonemes and syllables will be considered a false start (e.g. *av-avoiding* will be classified as a false start because its realisation is taken as: the first attempt, second attempt and final realisation).

Naming this topic is also another problematic issue. As can be seen, the terms repeat, repetition, and self-repetition are all widely used and often with the same intent. However, this paper will only talk about repeats, which Gráf (2017) describes in his paper. Gráf's definition differentiates *repeats* from *repetitions*, which involve the deliberate repetition of words or phrases for rhetorical or other purposes. He provides these to illustrate the difference:

1. a repeat: *"I mean the the play is really great."*
2. a repetition: *"The language really was very very nice."* (p. 67)

Distinguishing between these two is crucial as they serve different roles in spoken discourse. Repeats point to some cognitive processing limitations; in other words, they occur when the speaker struggles to plan or formulate their speech. They reflect the real-time demands of spontaneous speech production.[7] On the other hand, as mentioned before, repetition is an intentional tool speakers use to achieve certain goals, such as emphasising something or structuring discourse. Repeats are also often seen as markers of disfluency or hesitation and therefore might negatively impact the listener's fluency perception. In contrast, repetition indicates the speaker's intention and can improve clarity and the listener's comprehension.

2.3.2.6 Why repeats occur

Repeats occur because of cognitive and speech production challenges. Clark and Wasow (1998) explain that repeats often happen as part of the *commit-and-restore model*, where a

[6] They provide the example "*I, uh, I wouldn't be surprised at that.*" They state that „the speaker starts with "I", pauses with "uh", then repeats "I" before completing the phrase (p. 202).
[7] as the speaker is forced to "produce the language on the spot"

speaker prematurely commits to a word, suspends speech, and restarts with a repeat. "In the commit-and-restore model, repeats arise as speakers are trying to produce constituents, especially major ones such as noun phrases (NPs), verb phrases, prepositional phrases, clauses, sentences." (Clark & Wasow, 1998, p. 204) This process allows the speaker to "buy time" for planning and repair while keeping the other party still engaged. An example of this would be "I uhm I think…", where the speaker hesitates to form an idea but still manages to keep the listeners focused on them.

As mentioned in the chapter on false starts, repeats can also happen within a wide range of proficiency levels. Like false starts, they can happen to less advanced speakers who are simply searching for words, but also to more advanced ones, whose speech rate might be higher, for example. Consequently, they can let their "thoughts overtake their words, leading to repeats.

2.4 Language transfer

Second language acquisition is influenced by numerous factors, among which is language transfer. It is the process by which a speaker's L1 affects their L2 at various linguistic levels. When learning a new language, speakers often rely on the structures, pronunciation patterns, and speech habits they have already learned in their mother tongue. This influence of L1 can be either beneficial, because it can help learners see similarities between the two languages and therefore assist their learning process, or problematic, when the differences lead to errors or disfluencies (Odlin, 1989). Transfer happens naturally when learning a language, and it can influence our pronunciation and grammar, as well as when and how we, for example, hesitate.

Now that false starts and repeats have been examined in detail, it is necessary to consider how language transfer might explain why Czech speakers of English produce these disfluencies in their L2. Czech and English are both Indo-European languages; however, they differ significantly in their phonetics, syntax, and fluency strategies, which leads to observable instances of transfer. Examining how Czech speakers' L1 influences their speech in English helps determine whether the tendency to produce frequent false starts and repeats in their native language carries over into their L2.

2.4.1 Types of language transfer

Language transfer is generally divided into two main categories: positive transfer and negative transfer (also called interference) (Odlin, 1989).

Positive transfer occurs when the structures of L1 and L2 are similar, which makes language acquisition easier for learners. For example, Czech and English both share a subject-verb-object

(SVO) word order in basic declarative sentences (*He reads a book. – On čte knihu.*)[8]. Since the syntactic structure is similar, Czech speakers do not have significant difficulties with this. According to Odlin (1989), positive transfer can often lead to the speaker sounding more fluent, as learners can automatically apply familiar structures, which reduces cognitive load during speech production.

Negative transfer occurs when L1 and L2 differ, which leads to errors or unnatural speech patterns (Odlin, 1989; Verhoeven, 1989). Czech, unlike English, lacks the continuous aspect, which results in Czech speakers often avoiding or misusing progressive tenses (*She reads now* instead of *She is reading now*). This can also cause hesitation or pauses when speakers try to construct an unfamiliar verb form in real time (Belz et al., 2017). Belz et al. (2017) suggest that these pauses and errors originate from increased cognitive processing demands, as L2 speakers must adapt to structures that do not exist in their native language.

Both types of transfer contribute to disfluency production, but it is negative transfer that is most relevant to this study, as it can lead to increased occurrences of false starts and repeats (Tree, 1995; Belz et al., 2017). Studies on second-language fluency suggest that speakers have greater difficulty producing spontaneous speech when L1 and L2 have substantial grammatical differences, leading to more frequent pauses, repeats, and reformulations (Belz et al., 2017). These findings support the view that disruptions in fluency are not solely caused by low proficiency but can also result from the influence of L1 structures on L2 processing.

2.4.2 Transfer in phonetics, morphology, syntax and fluency

The way Czech speakers of English process and produce language is shaped mainly by their L1 phonological, syntactic, and fluency patterns. This influence affects grammar, vocabulary, pronunciation, speech rhythm, and disfluency patterns, all of which can contribute to false starts, pauses, and repeats. Studies on crosslinguistic influence show that transfer effects occur in spontaneous L2 speech, particularly when L1 and L2 differ significantly in phonological and syntactic structures (Odlin, 1989; Verhoeven, 1989).

2.4.2.7 Phonetic transfer

Czech stress patterns differ significantly from those in English, with Czech typically placing stress on the first syllable of a word regardless of length or morphology. In contrast, English varies stress based on lexical and grammatical function (Belz et al., 2017). This often leads to hesitation or mid-word pauses as Czech speakers attempt to recall and apply the correct

[8] Czech, of course, has a much more flexible word order than English, so the pronoun in the example above can be moved after the object of the sentence in a certain context, or even omitted entirely.

stress pattern (/dɪvə'lopmənt/ instead of /dɪˈveləpmənt/). In their study, Belz et al. (2017) noticed that advanced German learners of English often produce hesitations and self-corrections when dealing with English stress patterns, particularly in words where the stress placement is different from that of their native language. This is similar to the challenges faced by Czech speakers, as mentioned above, with the example /*dɪvə'lopmənt*/ instead of /*dɪˈveləpmənt*/. Czech also lacks certain English vowel distinctions, which can lead to self-corrections, prolonged hesitations, or false starts when speakers try to differentiate vowel length (Koponen & Riggenbach, 2000). Common problematic pairs for Czech learners include, for example, *sheep vs ship*, *pool vs pull*, and *beat vs bit*.

2.4.2.8 Morphological and syntactic transfer

Czech is a highly inflectional language, which allows it to have a relatively flexible word order. On the other hand, English relies on word order and auxiliary verbs. Morphological transfer, the influence of a speaker's L1 morphology on their L2, has often been underestimated in second-language acquisition research. However, Jarvis & Pavlenko (2008) argue that "the claim that overt inflectional morphology does not transfer is not only inaccurate but also obfuscates the fact that L2 users very frequently do exhibit CLI effects in their language comprehension and production" (p. 93). Studies have documented instances where L1 speakers apply native inflectional morphology to L2 words, such as Czech learners of Russian applying Czech inflectional endings on Russian words (*rabotnice* instead of *rabotnicy* = "workwomen") (Jarvis & Pavlenko, 2000, p. 92). Another example they listed was the case of Finnish and Swedish speakers of English[9], who transferred their spatial markers when using English prepositions. In their experiment, the participants were shown a scene from a film where the characters were sitting in the grass. As Jarvis and Pavlenko (2008) state, "The Finnish-speaking participants overwhelmingly described the protagonists as sitting *on* the grass, whereas the Swedish-speaking participants showed a preference for saying that they were sitting *in* the grass. The use of either preposition is acceptable in English depending on the length of the grass; the native English control group used *in* far more frequently than *on*, given that the grass in the film is quite high" (p. 93). These findings suggest that morphological transfer is an important aspect of crosslinguistic influence (language transfer).

[9] An interesting point to note is that Swedish is part of the Germanic language family (as well as English), whereas Finnish belongs to the Finno-Ugric languages. This, of course, causes Swedish to have "more in common "with English than Finnish.

Syntactic transfer (the influence of a speaker's L1 syntactic structures on their L2) has often been assumed to be minimal in second-language acquisition research. However, recent studies show that this assumption might be incorrect. As Jarvis and Pavlenko (2008) write, "Syntax, like morphology, has been widely assumed to be immune to CLI effects. […] Yet, this assumption is not warranted, and recent studies have documented ample instances of syntactic transfer in various types of data" (p. 96). A study by Zobl (1992) found that learners from different L1 backgrounds showed distinct patterns in their acceptance and rejection of ungrammatical sentences influenced by the syntactic norms of their first language (Jarvis & Pavlenko, 2008). Another example of this was as follows: "One example of this phenomenon was observed by Jarvis (2003) in a case study of a Finnish-English bilingual who rejected some grammatical L1 Finnish sentences that violated the less flexible word order constraints of her L2. That is, the relatively rigid SVO word order constraints of English appeared to influence her intuitions of well-formedness in her L1 Finnish" (Jarvis & Pavlenko, 2008, p. 97). These findings suggest that syntactic transfer is a significant factor in L2 acquisition, which helps both language comprehension and production.

2.4.2.9 Fluency and disfluencies transfer

Czech speakers often transfer speech planning strategies from L1 to L2 English, which can lead to disfluencies that reflect habits in the native language rather than deficiencies in the second language. For example, in English, a repeat is not necessarily perceived as hesitation. As Gráf (2017) argues, such repeats can serve as a supportive means to help the speaker control the conversation when planning the next utterance. When transferred to English, however, the same behaviour can be interpreted as a sign of lower fluency due to the different norms of what qualifies as 'fluent' speech in the two languages.

Although these repeats can be perceived negatively by listeners, researchers have repeatedly argued that they perform a crucial function in speech planning, especially for L2 learners. Götz (2013) argues that "these performance phenomena should not be evaluated negatively because they contribute to an impression of naturalness in speech" (p. 33). Disfluency features like repeats and false starts occur in predictable, systematic ways, not just as signs of breakdown. As she explains, "since performance phenomena are not used randomly but in a systematic way, i.e. at certain positions, and with certain functions [...] their usage enhances the speaker's productive fluency" (Götz, 2013, p. 42).

The key factor here is the pressure to plan. In spontaneous speech, all speakers need time to organise their thoughts, but in L2 speech, this requirement increases greatly. Götz (2013) notes that "there is high planning pressure especially in spontaneous speech and that it is

increased when a foreign language is being spoken. Naturally, more planning phases are necessary when speaking in the L2" (p. 34). As a result, L2 learners often rely on more than one repeat to maintain fluency. "One repeat at the beginning of an utterance is not usually enough even for advanced learners to cover their planning demands and in the majority of cases more performance phenomena follow after the repeat, most of them in the form of disruptive unfilled pauses" (Götz, 2013, p. 36). This implies that learners transfer speech production habits from their L1 context to their L2 context—not only in grammar or pronunciation but also in how they plan and produce speech in real time. Jarvis and Pavlenko (2008) refer to this as a conceptual or discourse-level transfer. They stress that "crosslinguistic influence may also affect learners' message construction processes, such as their selection and organisation of propositional content and their planning of speech acts" (p. 123). It also influences how learners organise their speech, including "the selection and sequencing of speech components, pause placement, and discourse structuring, even in cases where the speech is grammatically accurate" (Jarvis & Pavlenko, 2008, p. 126).

In this way, even the 'fluent' and accurate speech of a learner can have traces of L1. "Transfer need not be manifested solely in deviant or erroneous behaviour; it can also occur in subtle ways in fluent and accurate L2 performance" (Jarvis & Pavlenko, 2008, p. 121). What and how learners say is often influenced by the norms of conceptualising and organising discourse in L1: "What people say in a second language and how they say it may be influenced by the language-specific conceptualisations and discourse patterns of their L1" (Jarvis & Pavlenko, 2008, p. 123). This is especially relevant in the case of Czech speakers, whose L1 norms support flexible word order, a high tolerance for filled pauses and repeats, and pragmatic repetition at the beginning of utterances. When these strategies are transferred to English[10], the result can be perceived as non-fluent - even when the message is clear. As Götz (2013) puts it, "learners, no matter how advanced, will probably never wholly eliminate their higher planning pressure compared to speaking in their L1, but they could use the above strategies to make up for it in an elegant and nativelike way without even losing their turn" (p. 42).

All of this is evidence of a clear conclusion: fluency transfer from a speaker's first language into their second is not only common but something that researchers have observed frequently. As shown by both Götz (2013) and Jarvis and Pavlenko (2008), even very advanced speakers do not simply "switch off" their L1 habits when speaking a second language. Instead, they draw on strategies they already know - consciously or not - to manage real-time speech.

[10] where fluency and minimal overt planning are expected,

2.4.3 Research on the transfer of fluency and disfluency across languages

The idea that fluency (especially disfluencies such as pauses, false starts and repeats) is shaped not only by L2 proficiency but also by the habits of speakers in their first language has gained increasing support in recent research. Several empirical studies suggest that what we call "fluency" in a second language is often an extension or reflection of fluency behaviour in the speaker's native language. This section focuses on recent work that addresses the cross-linguistic transfer of fluency and disfluency between L1 and L2 and discusses how their results contribute to this research.

A key work closely related to the aims of this paper is Gráf (2017), who examined over 1,900 instances of repeats in 13 hours of English speech by Czech learners. Gráf notes that clause-initial repeats, such as "I I think that…" or "and and then we…" are not merely random hesitations, but rather phenomena that "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".[11] His data show that such repeats "are predominantly used at the beginnings of clauses or of nominal/prepositional phrases, where planning pressure is felt most acutely", which is a conclusion that supports the view that they perform a pragmatic rather than an erroneous function. An important point is that Gráf raises the issue of transfer explicitly and argues that "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" (p. 76). Gráf's research is important not only because it focuses in detail on types of disfluencies, such as clause repetition, but also because it examines the same combination of languages as this paper: Czech as L1 and English as L2. However, while studies focusing on Czech-English transfer are rare, several researchers have investigated fluency transfer in other language pairs.

One such study is Peltonen (2018), who studied the speech of 42 Finnish learners of English at two school levels. The analysis used both quantitative and qualitative tools to analyse speech samples in both languages. The results showed that "L1 fluency is an important factor in explaining L2 fluency and should be more widely acknowledged in L2 fluency research, assessment, and teaching" (Peltonen, 2018, p. 688). Especially relevant is her study of stalling

[11] Gráf defines repeats as "segments of speech which are involuntarily repeated in close proximity without adding any propositional content to the message" (p. 67), which is the definition also used in this work to define them. As mentioned before, he distinguishes them from intentional repetitions used for emphasis. His corpus analysis involved identifying these segments across clause types and comparing their distribution to those found in native speaker data.

mechanisms, including repeats and filled pauses, in both languages. She did not always find statistical correlations between L1 and L2 stalling, but Peltonen observed that "comparing L1 and L2 stalling mechanisms qualitatively helped to identify idiosyncratic patterns in their use" (p. 685)[12].

A more statistical approach was taken by Duran-Karaoz and Tavakoli (2020), who studied the relationship between fluency measures in Turkish (L1) and English (L2) using a large set of breakdown and repair metrics. They found moderate to strong positive correlations between the two languages in several areas: "number of mid-clause filled pauses ($r = 0.60$, $p < 0.001$), number of end-clause filled pauses ($r = 0.30$, $p < 0.048$), number of mid-clause silent pauses ($r = 0.34$, $p < 0.024$), and number of repair ($r = 0.45$, $p < 0.003$)" (Duran-Karaoz & Tavakoli, 2020, p. 684)[13]. These results strongly support the idea that fluency behaviour is not independent of different languages but is instead possibly shaped by consistent speech patterns.

Similar conclusions were reached in a longitudinal study by Derwing, Munro, Thomson, and Rossiter (2009). They followed two groups of recent immigrants to Canada: 16 speakers of Mandarin and 16 speakers of Slavic languages. They aimed to determine to what extent temporal features of L1 speech correlate with developing L2 fluency in English over time[14]. Participants did identical narrative tasks in both their native language and English at three time points over a year. The researchers observed significant correlations between L1 and L2 fluency measures (such as speech rate, number of pauses per second) during early stages of L2 exposure for both groups, but later, "only for the Slavic group" did those correlations remain strong. This led them to propose that the continued L1–L2 fluency relationship in the Slavic group "might be attributed to different amounts of exposure to English or to a closer relationship between Slavic languages and English than between Mandarin and English" (Derwing et al., 2009, p. 534). Despite being based in an immersion setting (Canada), the study remains relevant because participants were at an early stage of English acquisition, and the study directly compared performance in both languages. Moreover, the fact that only some learners kept cross-linguistic

---

[12] Her qualitative method involved selecting six speakers with frequent stalling phenomena in their L1 and comparing how they used similar mechanisms in English. This showed cases where the same hesitation strategy, such as repetition of conjunctions or fillers, was present across both languages, showing that fluency (or rather disfluency) was transferred from L1 to L2.

[13] Their methodology involved a series of monologue tasks in both Turkish and English. They focused especially on "mid-clause" hesitations, which suggest high planning pressure areas and may show transfer, like clause-initial repeats in Gráf.

[14] The "Slavic" group included Russian, Ukrainian, and Serbo-Croatian speakers, all of whom were recent immigrants living in Canada for two months at the first data collection point. Their narrative tasks were recorded in both their L1 and English, and analysed for pruned syllables per second, number of pauses per second, and overall speech rate.

fluency similarities while others did not give valuable context for this study's focus - exploring whether speaker-specific disfluency habits persist across language boundaries, particularly for closely related language pairs such as Czech and English. Although there were no native Czech speakers among this group of participants, the results are still relevant to this paper - according to their findings, speakers of all Slavic languages experienced language transfer of specific fluency strategies or disfluencies, and since Czech is a Slavic language, it is reasonable to assume that the results would not differ there either.

Another observation comes from de Jong, Groenhout, Schoonen, and Hulstijn (2015), whose work is frequently cited in recent fluency transfer studies (such as Peltonen or Duran-Karaoz and Tavakoli). Working with English and Turkish learners of Dutch, they found that "all correlations between L1 and L2 fluency measures were statistically significant, the strength ranging from r = 0.37 in mean syllable duration to r = 0.76 in silent pause (SP) duration" (as cited in Peltonen, 2018, p. 686)[15]. They argue that adjusting L2 fluency scores for L1 fluency provides a more accurate measurement of true L2 ability. This supports the core assumption of this paper that fluency and disfluency in English should be understood according to each speaker's Czech fluency profile.

Taken together, these studies show that the question posed by Gráf - "whether the use of repeats by L2 speakers mirrors their use of this strategy in their L1" (Gráf, 2017, p. 75) - is not only valid, but already being explored across different language pairs using a range of well-developed methodological frameworks. This research contributes to this growing field by focusing on false starts and repeats, which remain understudied compared to general measures such as pause frequency or speech rate. By focusing on false starts and repeats, this study contributes to the growing body of research on fluency transfer by asking whether disfluency patterns in Czech are carried over into English. If they are, this work examines how the word classes involved in these disfluencies (repeats specifically) differ between the two languages.

2.5 Research questions

This thesis, as previously described, focuses on disfluencies, repeats and false starts to be exact. The previous chapters dealt with defining the terms and presenting past research that addressed similar topics. As can be seen, apart from Gráf's (2017) research, there are not many studies on teachers' language, English, compared to their mother tongue, Czech (and thus L1-

[15] Their regression analysis found that L1 fluency correction improves the predictive power of L2 fluency measures such as syllable length, especially when determining vocabulary knowledge.

L2 transfer within fluency). For this reason, I have chosen the following three research questions, the answers to which will help me better understand the tendencies in Czech teachers' speech:

1) Do speakers who produce a higher number of false starts and repeats in their L1, Czech, also produce a higher number of these disfluencies in their L2, English?
2) Which word classes appear in one-word repeats in L1 Czech and L2 English, and how do their distributions differ across the two languages?
3) Do speakers produce longer disfluencies (in terms of the number of words) in their L2 English than in their L1 Czech?

3 Analysis – data and method

3.1 English Teacher Corpus (ETC)

For my research, I used the English Teacher Corpus project (hereafter referred to as ETC), which is a spoken corpus that documents the English of Czech teachers of English as a foreign language (EFL), along with a parallel subcorpus of native English teachers teaching in the Czech Republic. This is a unique project that focuses neither on traditional language learners nor on the language of teachers in lessons, but on their English outside the classroom. As the project's website states, "ETC represents a pioneering project as it is the first of its kind to compile a spoken learner corpus specifically focused on EFL teachers' language use" (English Teacher Corpus, 2025).

3.1.1 How ETC works and its background

The recordings in the corpus track both spontaneous and semi-spontaneous speech through a structured interview consisting of five parts. These include:

1. Task 1: a monologue in English, where the teacher reflects on some parts of their teaching experience of their choice;
2. Task 2: a dialogue in English, which takes the form of a semi-guided conversation on topics related to education or the previous task;
3. Task 3: a description of a sequence of pictures in English;
4. Task 4: reading a short text aloud in English;
5. Task 5: a dialogue in Czech (their L1), whose content is identical/similar to Task 1.[16]

This structure allows us to study how different tasks affect speech and compare L2 and L1 performance for the same speaker. Since the final section is in Czech, researchers can observe the differences and similarities in fluency, structure, or non-fluency between the two teachers' languages.

According to the corpus metadata (updated in June 2024), the Czech subcorpus includes 25 teachers (19 female, six male) who self-assessed their English proficiency as B2 (n=1), C1 (n=9), C1+ (n=5), and C2 (n=10). The native-speaker subcorpus includes 15 teachers (8 female, 7 male; mean age 32.6) from countries such as the UK, the US, and South Africa. In total, the recordings last 12.5 hours and contain 76,122 tokens in the Czech subcorpus and 31,898 tokens in the native speaker data. All interviews were transcribed using the Whisper AI tool, further

[16] In the interviews, the teachers were given a few points that served as suggestions of possible topics to discuss. These were topics such as 'what you enjoy about being a teacher and what you don't enjoy' and 'what you would like to change in education'. Of course, they were given the freedom to talk about anything else, provided that it was related to their profession/life as a teacher.

processed using corpus tools such as EXMARaLDA, which offers a time-aligned transcription, and finally were checked by the project leaders.

In addition to being a great source of semi-natural spoken English for analysis, the ETC was created with several clearly defined goals. These include supporting research into teacher language use, contributing to teacher training, and improving how teacher language proficiency is assessed. The authors also emphasise that the corpus is meant to give students in teacher training programmes practical experience with research and data collection. The project's website states: "To engage L2 teacher-trainees in the process of corpus compilation as part of their pre-service university teacher training, providing them with practical experience and insights into their future profession" (ETC, 2024). As a participant, I conducted three interviews myself, thus supporting the project and furthering my own experience.

The interviews were anonymised, and the participating teachers gave informed consent through a profile form that also collected metadata (such as previous education or residence abroad). The corpus covers not only language but also teachers' reflections on their work and profession, which is often related to their pupils, teaching problems, educational policy, or modern concepts in teaching, such as using artificial intelligence. This makes the data useful for both quantitative linguistic analysis and qualitative research focused on discourse. It is beneficial as it provides insights "into the professional perspectives and experiences of teachers" (ETC, 2024).

Looking ahead, the ETC is expected to be made available through the KonText interface of the Czech National Corpus. There are also plans to extend the project internationally, with partner universities in countries such as Poland, Germany, Austria, and Taiwan contributing comparable recordings. The project will eventually become the International English Teacher Corpus and later the International Corpus of Teacher Language, which will include recordings of non-native teachers of other foreign languages. This expanded version "will offer data for both linguistic and qualitative content analysis".

3.2 Research focus, use of the corpus and data

In my research, I focus on disfluency features in the speech of Czech EFL teachers - specifically, the number of false starts and repeats they produce in both their L1 (Czech) and L2 (English). For this purpose, I analyse recordings from Part 1 (the English monologue) and Part 5 (the Czech monologue) of the ETC interview. These two parts were chosen because they contain the most natural and spontaneous parts of the interview. Unlike reading or picture description tasks, these monologues allow speakers more freedom in what they say and how they say it and are more suitable for studying individual speech production patterns.

I aim to find out whether teachers who produce fewer dysfluencies in their native language (Czech) also tend to produce fewer dysfluencies in the L2 they are teaching (English). If such a pattern exists, it could support the idea of a transfer of fluency-related traits from L1, suggesting that fluency is not only determined by second language proficiency. However, it may also reflect individual speech habits or cognitive strategies that are common to both languages (as discussed in previous chapters on language transfer). In addition to analysing the total amount of disfluencies, I also examine which parts of speech are most frequently repeated during hesitations in each language. This will allow me to see whether hesitation behaviour in L1 and L2 targets similar types of words or shifts based on language structure. Finally, I examine whether disfluencies tend to be longer in English than in Czech, to assess whether speaking in a second language increases the complexity of hesitation.

3.3 Method

To be consistent in marking disfluencies, it was necessary to establish a set of rules that would help me measure them. Firstly, when highlighting all the disfluencies in learners' speeches, I had to not only review the transcripts of the interviews but also simultaneously listen to them. By only seeing the transcripts, misinterpretations of specific passages could occur, as utterances are not complete most of the time, and the tone of the speaker's voice often reveals whether a passage is a deliberate occurrence in speech or a disfluency. Below, I will describe specific cases where a set of rules proved to be not only important but also necessary.

3.3.1 The distinction between false starts, repeats and everything else

As mentioned previously, I focus only on repeats, not self-repetitions[17]. This type of disfluency is not apparent simply from the transcript itself. Therefore, a recording is necessary to decide whether the utterance of two or more repeated words was intentional (self-repetition) or not (repeat). This also applies to the question of whether the speaker is self-correcting or if it is a FS. Again, the recordings either show that they only "stuttered", and therefore, it cannot be marked as a FS, or they paused slightly on the word/phrase, creating one.

Another rule I decided on when finding FS and R in transcripts was that I do not include filler words and sounds as words. As described above, disfluencies occur in speech for many reasons, whether it is nervousness or 'buying time'. I have observed that while some people use sounds such as *uhm, ehm, em*, or *er* when searching for words, others use filler words or phrases

[17] As Gráf (2017) distinguishes in his study on disfluencies, and as was previously mentioned in the chapter Repeats – Definition and naming them

instead. That is why when marking disfluencies, I have decided not to count as words the following list of words/phrases:

- in recordings in L2 (English): like, you know, okay
- in recording in L1 (Czech): vlastně, jako, jakoby, prostě, vlastně jakoby.[18]

To illustrate this, a repeat would be counted as one in all the following cases:

(1) *what I really enjoy enjoy is singing;*

(2) *what I really enjoy (er) enjoy is singing;*

(3) *what I really enjoy you know enjoy is singing.*

And a false start in the following ones:

(4) *what I don't what I really enjoy is singing;*

(5) *what I don't (uhm) what I really enjoy is singing;*

(6) *what I don't you know what I really enjoy is singing.*

3.3.2 Marking repeats and false starts

Firstly, all speakers were assigned a speaker ID to stay anonymous, which is why I will be commenting on their interviews by using these codes (such as CZ001, CZ024). In this research, FS and R are recorded using the labelling presented in Gráf (2017).[19] Individual disfluencies are recorded in the transcripts in the following ways:

(7) <R_1_2_P> *I I just thought of it last week*[20] (CZ010)

(8) *sometimes* <R_1_2_X > *it's it's like (eh) heavy responsibility on my shoulders* [21] (CZ002)

(9) <R_2_2> *she's just she's just walking around*[22] (CZ007)

(10) *it turned out that I* <FS_1> *would wouldn't stay in in Brno* (CZ010)

(11) <FS_3> *the textbook chan= the textbooks change* (CZ007)

In terms of the marking of the word classes in both languages, I have kept the same marking as Gráf:[23]

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

[18] Of course, all of these must be used as filler words to not be counted as a part of the speech.
[19] As this study focuses specifically on false starts and repeats, its labelling is well suited to the purpose of my paper.
[20] <R_1_2_P> = twice occurring preposition (= a repeat of a preposition which happened twice)
[21] <R_1_2_X> = twice occurring contraction (= a repeat of a contraction which happened twice)
[22] <R_2_2> = twice occurring repeat of two words
[23] "Meaning of the codes in the fourth position of the tags" (Gráf, 2017).

particle; J - adjective; N - noun; O - other; P - pronoun; R - rhetorical; V - verb; W - wh-word; X – contraction.[24]

To verify the correct word class of each word, I used the online version of *Cambridge Dictionary*[25] for words from task 1 (monologue in English) of *Slovník spisovného jazyka českého*[26] for task 5 (monologue in Czech).

The classification of disfluencies as either R or FS was based not only on the formal characteristics of the utterance but also on the speaker's assumed intention and the surrounding context. Unlike relying only on structural criteria, I chose to prioritise the interactional and cognitive function of the disfluency - that is, how it appeared within the spontaneous flow of speech. The tagging system I used was chosen to reflect this approach as clearly and consistently as possible across both tasks 1 and 5 (English and Czech monologues).

Repeats were defined as the immediate recurrence of a word or phrase without a clear shift in structure or meaning. These typically occurred during moments of hesitation, self-monitoring, or planning. In my tagging, repeats were included only when they interrupted the progression of the utterance or signalled a delay in formulation. I did not include stylistic or emphatic repetitions, nor repetitions that occurred as part of fixed expressions or rhythmic sequences.[27]

(12) <R_1_2_P> ***I I*** *just thought of it last week* (CZ002)

(13) <R_2_2> ***she's just she's just*** *walking around* (CZ007)

(14) *so I decided to:* <R_1_2_V> ***obtain*** *(er)* ***obtain*** *a license or (eh) qualification* (CZ025)

However, I did not count repeated fillers like *uh uh* or *erm erm* as repeats. These were treated as a separate class of disfluency and were not included in the FS/R analysis. Similarly, conjunctions that were repeated at the beginnings of utterances, especially when they were clearly interactional ("so, so what I wanted to say…"), were not tagged unless they disrupted the structure or delayed the message.

False starts were marked when the speaker initiated a phrase or clause and then abandoned it, which often led to them starting again with a different structure. What mattered in my decision-making was not only whether the word or phrase was repeated or interrupted, but whether the speaker was shifting course, which thus suggests some cognitive processing, repair, or reformulation. I identified two major patterns in my dataset that occurred repeatedly:

---

[24] It is important to note that contractions, in terms of counting the number of words repeated, are marked as 1: <R_1_2_X> *it's it's*.

[25] https://dictionary.cambridge.org/

[26] https://ssjc.ujc.cas.cz/

[27] as explained in the chapter about repeats ('Definition and naming them')

Unfinished false starts

In these cases, the speaker began a word or phrase and abandoned it midway. These were marked with an equals sign (=) and reflected moments where the speaker failed to retrieve the full lexical form or restarted before completing the word.[28]

(15) *I work* <FS_1> ***ma= mostly*** *with kids or mainly with kids* (CZ025)

(16) *aby si to prostě člověk <FS_1>* ***zku= zkusil*** (CZ005)

I chose to mark these not because they represented clear lexical errors, but because they signalled cognitive disfluency (a moment where the speaker's planning failed or where they hesitated). Even if the unfinished item was brief or phonologically small, it reflected a disruption.

Finished false starts

These included complete words or phrases that were then abandoned or restructured. I marked them when the speaker obviously changed the direction of the utterance or started it again, usually with a different structure or focus.

(17) *I'm struggling* <FS_1> ***to with*** *(er) the teaching of English* (CZ006)

(18) *aby zůstaly nějaký nároky ale* <FS_3> ***aby se to aby se*** *prostě* ***osekaly*** *ta fakta* (CZ006)

In both examples, the speaker initiates a grammatical structure but then abandons or restructures it, which indicates a shift in the intended direction of the utterance. In (17), the speaker begins with “to” but then changes course to “with,” suggesting a momentary hesitation. Similarly, in (18), the initial clause “aby zůstaly nějaký nároky ale aby se to” is interrupted and restructured into “aby se prostě osekaly ta fakta,” which reflects some redirection of thought.

Overlapping repeats and false starts

Occasionally, a false start and a repeat occur within the same segment of speech. These are cases where a speaker begins an utterance, interrupts themselves, and then restarts with the same word or phrase, resulting in a repeat within or following a false start. In the following example, the speaker begins to say "secondary school", but stops after the initial /s/ sound, then starts again with "secondary", which is immediately repeated:

(19) *the first idea came in a* <FS_1> ***s=*** <R_1_2_J> ***secondary secondary*** *school* (CZ007)

[28] The only instances where the equals sign was not used were in the case of task 5 (in Czech monologues). When the speaker produced a false start starting with *ne-*, the transcription program detected it as a full lexical word. Therefore, it did not use the same marking as with other unfinished false starts. This can be seen in the following example: *i když sem třeba* <FS_1> ***ne neudělala*** *úplně všechno* (CZ004).

Here, the false start is marked at the point of interruption (“s=”), indicating the speaker’s initial failed attempt. The restart, however, contains a repeat of “secondary”, which shows both the hesitation and the speaker's attempt to stay fluent.

The following instance (20) gives an example of a case where a false start is immediately followed by a repeat that continues the utterance. In (20), the speaker begins to say "pomáhá", abandons the structure, and then continues with a restarted phrase that begins with a repeated personal pronoun:

(20) *protože mi to* <FS_1> ***pomáhá*** <R_1_2_P> ***já já*** *sem ráda připravená* (CZ002)

The false start "pomáhá" appears as an initial attempt at the clause, but it is not completed or integrated syntactically with what follows. Instead, the speaker restarts the idea from a different point, and in doing so, repeats the pronoun "já". This results in a layered disfluency where the interruption (FS) is immediately followed by a repeat (R) as the speaker reformulates the utterance.

In my dataset, overlaps between false starts and repeats occurred more frequently in Task 1 (the English monologue) than in Task 5 (the Czech one). This difference may reflect higher cognitive demands or lower automaticity in L2 production, leading speakers to revise and repeat more often within a single stretch of speech. Regardless of language, though, these combined disfluencies needed to be tagged individually in order to maintain consistency and accuracy in the annotation process.

3.3.3 Counting the words in a recording

Unlike determining what is and is not a disfluency, when counting words in individual interviews, filler words (such as ‘you know’ or ‘like’) must be included.[29] This is because the word count is meant to reflect the speaker’s complete verbal output, not just the content-carrying words. Leaving filler words out would artificially lower the word total and risk exaggerating the relative frequency of disfluencies, especially in more hesitant or nervous speakers. Contractions are also counted as one word here, as they represent a single grammatical unit spoken as a single lexical item.

In the recordings, any speech produced by the interviewer (at the start of the interview or during it) was excluded from the analysis, as it does not reflect the language use of the teachers themselves. Additionally, many interviewees began their answers by reading the task prompt aloud before responding. This portion was deliberately excluded from the word count,

[29] For this reason, when counting words, a definition of a word would be “a sequence of letters enclosed by spaces”.

as it does not represent spontaneous speech, but rather the reproduction of a pre-written instruction. Including it would not only increase the word count, but the speaker's actual production choices could also be misinterpreted, such as lexical selection, fluency, or hesitation patterns. Since the goal of this study is to analyse natural, unrehearsed language in spontaneous monologues, only the language content that followed the reading of the prompt was considered. Similarly, in Task 5 (in L1, Czech), if a speaker began their monologue in English before switching to Czech, the English portion was excluded, because it does not contribute to the intended L1 dataset.

3.3.4 Calculations and tests

After highlighting all separate disfluencies within the transcripts, they all needed to be counted for any further research to be possible. This proved to be one of the most challenging parts of the thesis, as the results had to be carefully gone through many times with repeated listening to the recordings and subsequent corrections (if necessary).

To do this more efficiently, I used Python to automatically extract each marked inconsistency into an Excel table. This script allowed me not only to list the disfluency and its tag (e.g., <FS_1>, <R_1_2_P>), but also to include the surrounding context: specifically, the five words preceding the disfluency and the seven words following it. Filler terms such as *er*, *uh*, or *erm* were excluded from this count. These boundaries were chosen based on the most frequent tags in my dataset, which were typically <R_1_2> and <FS_1>. In these cases, the disfluent part itself often consisted of one or two words, meaning that the pattern "five words before - disfluency - five words after" could be effectively captured by using seven words after the marker. These concordance lines were then manually reviewed to ensure the immediate context made sense, adjusting the number of surrounding words if necessary to avoid broken structures or incomplete sentences. I then highlighted the disfluency itself and recorded its individual details.[30]

After annotating each speaker's disfluencies, I calculated their relative frequency per hundred words (phw) using a basic formula in Excel.[31] This step ensured that speakers with longer monologues were not overrepresented in the final analysis and allowed comparison between speakers regardless of transcript length.

To test whether speakers who produced more disfluencies in English (L2) also tended to produce more disfluencies in Czech (L1), I used Lancaster University's online *log-likelihood*

[30] for example, whether it was a repeat or false start, or how many words were repeated or restarted.
[31] = (amount of a particular disfluency type within that speaker/their word count)*100

calculator (UCREL Log-likelihood Wizard)[32]. This test determines whether the observed difference in disfluency counts between the two corpora is statistically meaningful or likely to have occurred by chance.

To study whether individual speakers produced disfluencies differently in Czech (L1) and English (L2), I used a paired t-test[33]. The paired t-test is designed to compare two related sets of values, in this case, each speaker's use of disfluencies in both languages, to determine whether their differences are statistically significant. What makes this method well-suited for this research is that it accounts for within-speaker variation, making it possible to assess whether disfluency patterns in L1 and L2 differ consistently across participants, rather than just in overall totals. For the analysis, I used relative frequencies of disfluencies per 100 words instead of raw counts. This was necessary because many speakers produced more words in one language than the other, for instance, speaking more extensively in Czech than English. If raw frequencies were used, higher disfluency counts might simply reflect longer speech rather than an actual difference in fluency. Relative frequencies enable a fair comparison because they show how often disfluencies occur in relation to the amount of speech produced. This step normalises the data and helps ensure that the statistical results reflect differences in fluency behaviour, not just differences in speech length. For each comparison (false starts, repeats, total disfluencies, and disfluency length), I entered two columns of speaker-level relative values (L1 and L2) into the calculator. The output included the mean difference, standard deviation, confidence interval, and p-value, which together indicate the size, direction, and statistical reliability of the observed difference between the two languages.

[32] https://ucrel.lancs.ac.uk/llwizard.html
[33] https://www.statskingdom.com/paired-t-test-calculator.html

4 Results

This chapter discusses the results of the study, covering both quantitative findings and selected qualitative observations. The analysis begins with a comparison of overall disfluency frequencies in Czech and English, followed by a more detailed examination of how specific disfluency types, false starts and repeats, behave across the speakers' L1 and L2. The lexical structure of one-word repeats is then examined to determine which parts of speech are most frequently repeated. Lastly, the average length of disfluencies (measured by the number of words involved) is compared between L1 and L2 to see whether speaking a second language leads to longer or more complex hesitations.

4.1 Comparison of total disfluency frequency in L1 and L2

Across the two Tasks, there were 547 disfluencies in total, of which 296 were in Task 1 (L2, English monologue) and 251 were in Task 5 (L1, Czech monologue). These data suggest that speakers produce fewer disfluencies in their native language. Taking into account that the total number of words from Task 1 and Task 5 differs further supports the observation that the total number of words produced in Task 5 is significantly higher than in Task 1.

To address this, the *relative frequency of disfluencies per 100 words* was calculated for each speaker. These values are presented in the table below, sorted by speaker and language:

Speaker_ID	Disfl_phw_E	Disfl_phw_CZ
CZ001	0.93	1.35
CZ002	2.51	0.85
CZ003	2.67	2.46
CZ004	1.70	2.14
CZ005	1.44	1.04
CZ006	3.61	1.99
CZ007	2.30	0.15
CZ008	1.44	2.19
CZ009	5.93	3.62
CZ010	4.72	2.86
CZ011	4.35	0.99
CZ012	0	0.16
CZ013	2.77	1.89
CZ014	3.27	1.57
CZ015	0.69	0.72
CZ016	0.51	0.32
CZ017	1.20	2.23

CZ018	1.41	0.88
CZ019	5.05	1.31
CZ020	1.46	0.77
CZ021	2.13	1.78
CZ022	3.73	2.88
CZ023	0.84	0.81
CZ024	0.56	1.01
CZ025	1.27	0.92

Table 1. *Relative frequencies of total disfluencies per speaker in L2 (English) and L1 (Czech)*

The next step was to evaluate whether the observed difference in total disfluency frequency was statistically meaningful at the corpus level. A log-likelihood test was conducted using raw disfluency frequencies and the total number of words across all interviews in Czech and English, respectively. The test was carried out using Lancaster University's online log-likelihood calculator. The resulting log-likelihood score was 30.50, exceeding the standard significance threshold of 3.84 ($p < 0.05$) and confirming that the difference in disfluency frequency between L1 and L2 is *statistically significant*. However, this corpus-level difference does not appear consistently at the individual level. When calculating the log-likelihood separately for each speaker, only a few speakers (e.g., CZ019, CZ007, CZ011) reached the critical threshold of 3.84.

This suggests that although there is a general trend in the dataset, the behaviour of individual speakers varies considerably. In other words, the overall result may be influenced by a subgroup of speakers who show substantial differences between L1 and L2, while other speakers show little or no differences.

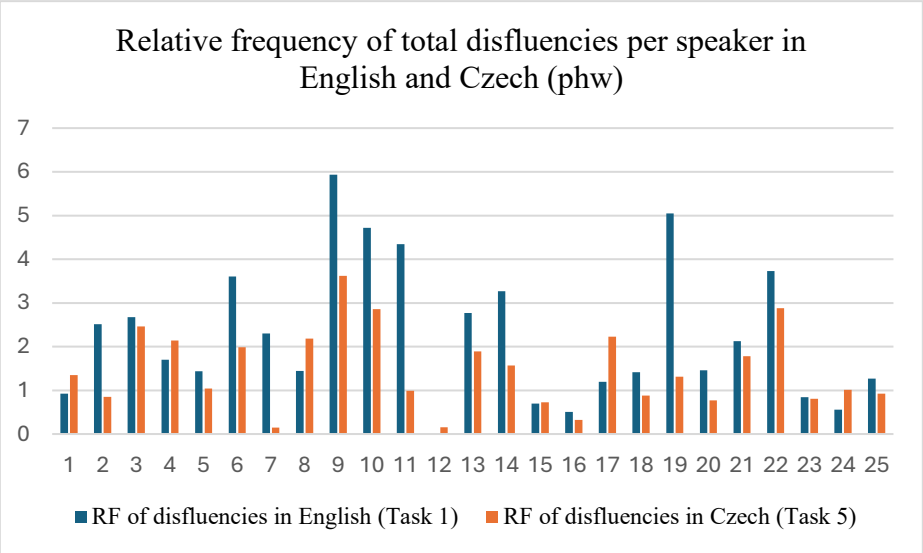

Figure 1. *Relative frequency of total disfluencies per speaker in English and Czech. Each pair of bars represents one speaker's disfluency rate phw in their L2 (English) and L1 (Czech).*

A *paired t-test* was then used to compare the relative frequency of total disfluencies per 100 words between the two languages. The significance level was set at $\alpha = 0.05$. Any result with a p-value below this threshold was interpreted as statistically significant. Speakers produced, on average, 2.3 disfluencies per 100 words in English (SD = 1.60) and 1.5 in Czech (SD = 0.90), $t(24) = 3.20$, $p = 0.004$. This result is *statistically significant* and confirms the difference observed at the corpus level, which suggests that speakers are overall less fluent in their L2. Unlike the log-likelihood test, which evaluated the distribution of disfluencies based on raw frequency, the paired t-test specifically compares variations within a single speaker. In this case, it is not only the number of disfluencies produced that matters, but the fact that most speakers tended to produce more disfluencies in English than in Czech. This supports earlier observations that only a few speakers showed a disproportionate number of disfluencies and hints that even though the difference at the corpus level is partly due to some outliers, the overall behaviour of the speakers is going in the same direction, just with different intensities.

In the table below, the mean relative frequencies of disfluencies per 100 words are presented for each disfluency type, alongside the corresponding *t*- and *p*-values. These results are important because they show whether the pattern observed in total disfluencies also holds true for the two main disfluency types: false starts and repeats. While the overall difference between Czech and English was statistically significant ($p = 0.004$), looking at the differences by type seems more nuanced. Repeats show a significant difference ($p = 0.037$), which suggests that speakers tend to repeat more in English than in Czech. In contrast, the difference in false starts is not statistically significant ($p = 0.065$), though it is just above the conventional threshold (α

= 0.05) and may still indicate a meaningful trend. A more detailed look at each disfluency type and how they behave individually across the two languages will be discussed in the next chapter.

Measure	mean (EN)	mean (CZ)	t	p
Total disfluencies (RF)	2.26	1.48	3.18	0.004
False starts (RF)	1.86	1.16	1.93	0.065
Repeats (RF)	1.29	0.87	2.20	0.037

Table 2. *Mean relative frequencies of disfluency types and t-test results (L1 vs. L2)*

4.2 Comparison of FS and R between L1 and L2

As disfluencies have been discussed as a whole so far, this chapter will address them separately, by comparing false starts and repeats individually. Not all disfluencies behave in the same way, and some played a bigger role in the difference between Czech and English than others. This section focuses separately on each type to show how they differ across the two languages. For each, I compared the number of occurrences and used paired t-tests to test the differences.

4.2.1 False starts – comparison of L1 and L2

The first type of disfluency analysed separately was the false start. As with total disfluency frequency, the number of false starts per 100 words (phw) was calculated for each speaker in both Czech and English. These relative frequencies are presented in Table 3 below. In addition, Figure 2 visualises each speaker's production of false starts across the two languages.

Speaker ID	FS_E_phw	FS_CZ_phw
CZ001	0.37	1.12
CZ002	1.14	0.57
CZ003	0.95	0.92
CZ004	1.02	0.91
CZ005	0.77	0.70
CZ006	1.14	1.05
CZ007	1.29	0.15
CZ008	1.08	0.82
CZ009	1.84	1.64
CZ010	1.10	1.19
CZ011	2.70	0.20
CZ012	0	0.16

CZ013	0	0.29
CZ014	1.22	1.26
CZ015	0.69	0.36
CZ016	0.34	0.16
CZ017	0.80	0.45
CZ018	0.35	0.50
CZ019	2.16	0.36
CZ020	0.97	0.58
CZ021	0	0.48
CZ022	2.19	0.41
CZ023	0.67	0.35
CZ024	0.37	0.20
CZ025	1.01	0.26

Table 3. *Relative frequency of FS per speaker in English and Czech (phw)*

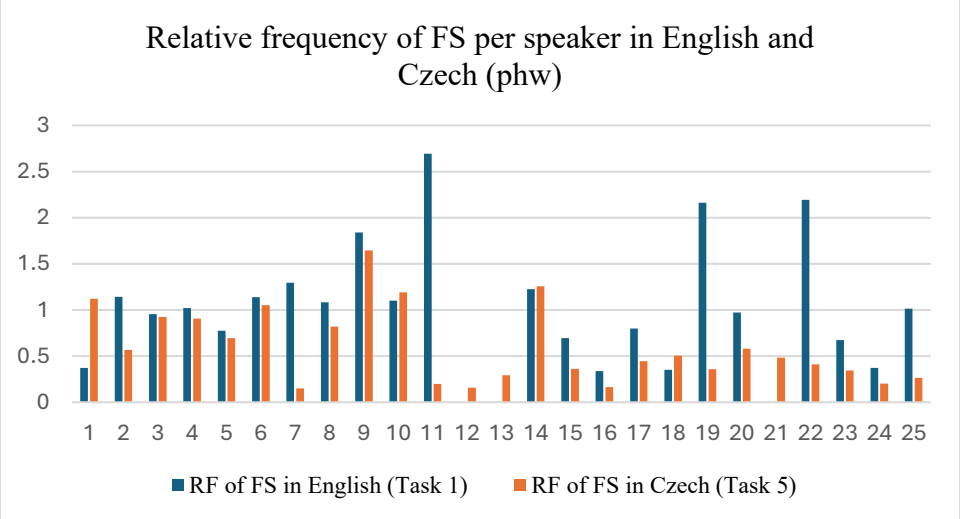

Figure 2. *Relative frequency of FS per speaker in English and Czech (phw)*

On average, speakers produced one false start per 100 words in Czech and 0.60 false starts per 100 words in English. Results of the paired t-test indicated that there is a significant small difference between Czech and English, $t(24) = 2.5$, $p = 0.021$. Speakers tended to produce more false starts in their native language than in their second language, although the difference is relatively small. Individual speaker results, as seen in Figure 2, reveal considerable variation, with some speakers producing more false starts in English, others showing minimal difference between the two languages, and several producing more in their L1. This suggests that the occurrence of false starts across languages is highly individual and does not indicate a consistent transfer effect at the group level. The data also suggests that speakers tended to initiate false starts similarly in both languages, most often after short introductory words such as pronouns

("I," "we") or conjunctions ("and," "but"). There was no clear evidence that false starts in English were systematically longer or more complex than in Czech, based on the data from the T1 and T5 tasks.

4.2.2 Repeats – comparison of L1 and L2

The second type of disfluency analysed separately was repeat. As with false starts, the total number of repeats was counted for each speaker in Czech and English. The results for individual speakers are shown in Table 4 below. Figure 3 gives a visual comparison of the two languages.

Speaker ID	R_E	R_CZ
CZ001	0.56	0.22
CZ002	1.37	0.28
CZ003	1.72	1.54
CZ004	0.68	1.23
CZ005	0.66	0.35
CZ006	2.47	0.93
CZ007	1.01	0
CZ008	0.36	1.37
CZ009	4.09	1.97
CZ010	3.62	1.67
CZ011	1.66	0.79
CZ012	0	0
CZ013	2.77	1.60
CZ014	2.04	0.31
CZ015	0	0.36
CZ016	0.17	0.16
CZ017	0.40	1.78
CZ018	1.06	0.38
CZ019	2.88	0.95
CZ020	0.49	0.19
CZ021	2.13	1.29
CZ022	1.54	2.47
CZ023	0.17	0.46
CZ024	0.19	0.81
CZ025	0.25	0.66

Table 4. *Relative frequency of R per speaker in English and Czech (phw)*

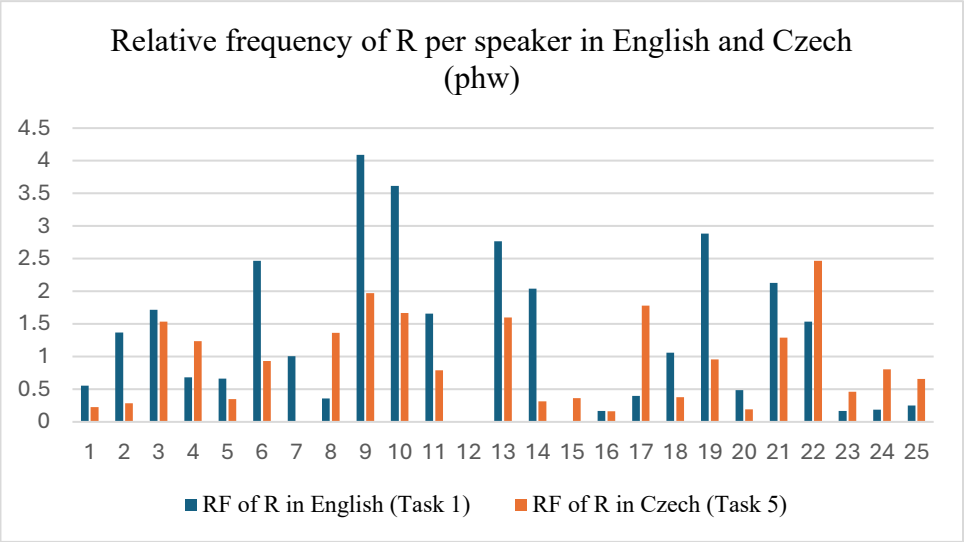

Figure 3. *Relative frequency of R per speaker in English and Czech (phw)*

On average, speakers produced 6.70 repeats in English and 5.00 repeats in Czech. A paired t-test indicated that this difference was not statistically significant, $t(24) = 1.70$, $p = 0.103$. Nevertheless, the difference in means points clearly towards a systematic tendency for speakers to produce more repeats in their L2, English. Given the relatively small sample (25 speakers), it is likely that the corpus simply does not have sufficient statistics to confirm this tendency at a conventional level of significance ($\alpha = 0.05$).

A closer look at Figure 3 suggests that most speakers repeated more words in English than in Czech. This pattern strongly supports the hypothesis that disfluency behaviour transfers from L1 to L2, particularly when it comes to repeating themselves. Only a few speakers showed a different pattern, using a similar number of repeats or slightly more in Czech. These cases seem to be isolated exceptions that do not reflect the general behaviour of the group. The overall distribution across speakers points very clearly towards a transfer from L1 to L2.

When analysing individual repeated items from tasks T1 and T5, there is further support for L1-L2 transfer. In both languages, speakers most frequently repeated short grammatical words such as pronouns ("I," "he," "it") and conjunctions ("and," "but," "ale"). For example, "I I" (CZ010) and "and and" (CZ003, CZ005) were common repeats during the English monologues, while items similar to "já já jsem" (CZ014) and "a a potom" (CZ020) appeared frequently in the Czech data. These repeats occurred mainly at the beginning of clauses or after the speakers hesitated, which suggests that they rely on similar hesitation strategies in both their native and second language. While the structure of repeated elements stayed consistent across both languages, the frequency with which these repeats occurred was noticeably higher in English for most speakers. This supports the idea that although the basic strategies of speech

disfluencies in speakers are based on speech habits in their first language, cognitive demands when speaking in a second language cause these strategies to occur more frequently.

In summary, although the statistical test did not reach the conventional threshold for significance, the observed trends in the data strongly suggest the presence of L1-L2 transfer in the use of repeats. Repeated patterns and consistency among most speakers suggest that this type of disfluency reflects basic speech production strategies that speakers carry over from their first language when dealing with the higher cognitive demands of the second language.

4.2.3 Comparing the behaviour of false starts and repeats

Several observations emerged when examining false starts and repeats in Czech and English. Although both types of disfluencies tended to occur more frequently in English, repeats followed a much more regular and predictable pattern than false starts did. False starts showed significant individual differences - some speakers even had fewer false starts in English than in Czech - and the t-test confirmed that this difference is not statistically significant. Repeats, on the other hand, were much clearer: most speakers repeated more words in English, and what they repeated, especially short grammatical words, remained surprisingly consistent across both languages.

Something I noticed while going through the data was how differently these disfluencies fit into the structure of a sentence. Repeats almost always showed up at the beginnings of clauses, usually involving words like pronouns ("I," "he") or conjunctions ("and," "but"). It felt almost automatic, as if when speakers hesitate, they grab onto something small and familiar to keep their speech moving forward. False starts, by contrast, were less predictable. Sometimes they interrupted whole stretches of speech, and sometimes they led speakers to completely rethink the structure of what they were about to say.

Repeats also seem to hint at a lighter cognitive load. Speakers simply repeat the word they were already planning to say, giving themselves some more planning time without having to change the sentence. False starts, on the other hand, often mean a bigger decision: something wasn't quite right with what was about to be said, and so the speaker had to make a real-time adjustment. It makes sense, then, that repeats (the simpler and more automatic strategy) would show a clearer and more consistent transfer from L1 to L2, while false starts would remain messier and more speaker-dependent.

Overall, although both types of disfluencies show how speakers cope with the pressures of second language production, repeats provide the strongest indications of transfer.

4.3 Lexical characteristics of repeats

After looking at the overall frequency of disfluencies and how false starts and repeats behave across languages, I wanted to take a closer look at what precisely the speakers were repeating. Numbers alone do not tell the whole story. Repeats are not just random hesitations - they reveal where the speaker pauses for a moment while planning, and which parts of the sentence are so automated to the speaker that they repeat them without thinking.

In this chapter, I focus specifically on one-word repeats - the moments when a single word is repeated right before the speaker continues their sentence. By examining the types of words that speakers repeat in Czech and English, I aim to better understand whether hesitation patterns are rooted in first-language habits and carried into second-language speech. This analysis addresses Research Question 2 (RQ2): *Which word classes appear most often in one-word repeats in L1 (Czech) and L2 (English)?*

For each language, I compared the distribution of word classes (pronouns, conjunctions, verbs, and others) to see whether similar patterns appear. The results give a clearer picture of how hesitation and planning work in spontaneous speech, and how much of this behaviour carries over from L1 to L2.

4.3.1 Word classes in Czech repeats

I started by analysing one-word repeats produced during the Czech monologues. Table 5 below shows how often different word classes were repeated, and Figure 4 visualises their proportions. As the table suggests, certain types of words were much more likely to be repeated than others. Pronouns and conjunctions made up a large part of all one-word repeats, which already hints at some interesting patterns in how speakers manage hesitation in their native language.

I first looked at the word classes involved in one-word repeats produced during the Czech monologues. Table 5 shows the frequency of different word classes being repeated, and Figure 4 visualises their proportions. Even at first glance, some interesting patterns become clear. Certain types of words were repeated much more often than others, suggesting that hesitation tends to cluster around specific parts of a sentence.

Conjunctions ("a," "ale") made up by far the largest share of one-word repeats, accounting for almost 40% of all cases. Pronouns ("já," "on," "to") followed, representing around 20%. Adverbs ("pak," "tak," "už") came next, with 15.6% of the total, followed by prepositions and verbs. The dominance of conjunctions and pronouns seems to reflect the natural structure of Czech monologues, where speakers frequently need to link clauses together or momentarily hesitate when retrieving a subject or restarting a phrase.

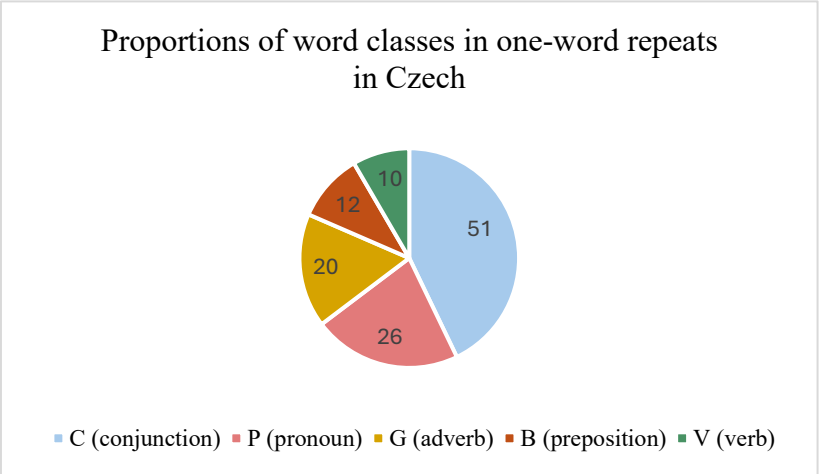

Figure 4. *Proportions of word classes in one-word repeats in Czech. For specific counts and percentages, see Table 5.*

What became particularly noticeable was the concentration of repeats around small grammatical "function" words. Content words, such as nouns or adjectives, barely appeared among the repeated items. This suggests that when speakers hesitated in Czech, they often did so while dealing with grammatical words, not lexical ones. It suggests that hesitation is something closely tied to planning the structure of speech, rather than searching for content creation. However, it is also possible that speakers repeat function words while planning the upcoming lexical material, particularly when retrieving content words proves more difficult. A detailed analysis of the words following each hesitation would be necessary to confirm this, but such an analysis was beyond the scope of the present study.

Word Class	Count	Percentage
C (conjunction)	51	39.80%
P (pronoun)	26	20.30%
G (adverb)	20	15.60%
B (preposition)	12	9.40%
V (verb)	10	7.80%

Table 5. *Distribution of word classes in one-word repeats in Czech. Only conjunctions, pronouns, adverbs, prepositions, and verbs were observed.*

The concentration of hesitation around these small grammatical elements reveals the role of clause structure in spontaneous speech. Speakers typically did not hesitate over heavy content words or complex ideas; instead, they paused at the connective points—the moments where one clause links to the next or where a new subject is introduced. This pattern suggests that when planning speech under time pressure, speakers are not only retrieving vocabulary but are actively organising the grammatical framework that holds their sentences together.

Although only five word classes appeared in the one-word repeats from the Czech monologues, their distribution already indicates the underlying processes. The strong prevalence of conjunctions, in particular, shows the importance of smooth transitions between ideas in maintaining the flow of speech and how sensitive these points are when the speaker hesitates.

4.3.2 Word classes in English repeats

After analysing the word classes in Czech one-word repeats, I moved on to the English monologues. Table 6 and Figure 5 show how the distribution of repeated word classes shifted when the speakers were speaking in their second language. Although some patterns remained similar, the differences in proportions suggest that L2 speech may subtly influence which parts of a sentence speakers rely on during moments of hesitation.

Pronouns were by far the most frequently repeated word class in English, accounting for 45% of all one-word repeats. This is a noticeable shift compared to Czech, where conjunctions dominated. Conjunctions and prepositions followed next in English, each making up 7.4% of the repeated items, with definite articles (8.1%) and contractions (6.0%) also appearing among the most repeated word classes. The prominence of pronouns in English may reflect the more rigid syntactic requirements of English sentences, where subject pronouns are almost always needed, unlike in Czech. However, a fuller discussion of this point will follow later.

Word Class	Count	Percentage
P (pronoun)	67	45.00%
Ad (definite Article)	12	8.10%
C (conjunction)	11	7.40%
B (preposition)	11	7.40%
X (contraction)	9	6.00%

Table 6. *Distribution of word classes in one-word repeats in English. Only pronouns, definite articles, conjunctions, prepositions and contractions were observed.*

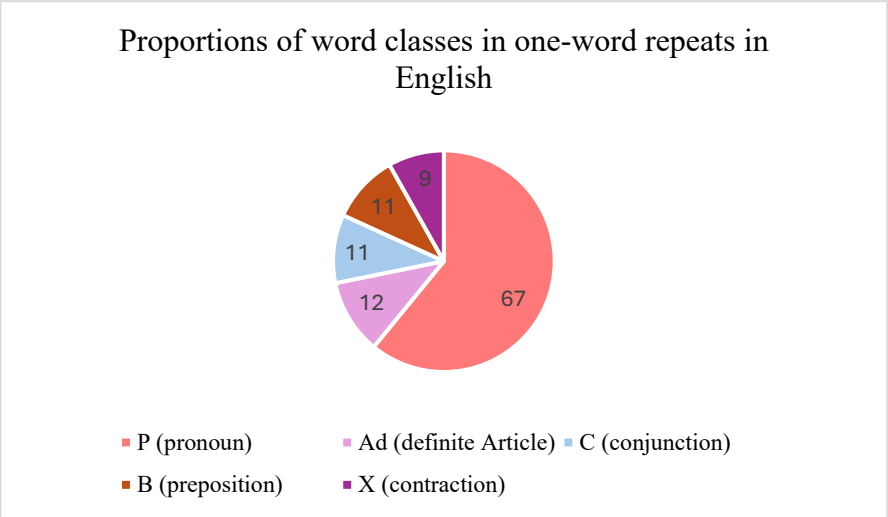

Figure 5. *Proportions of word classes in one-word repeats in English. For specific counts and percentages, see Table 6.*

The overall picture shows that although hesitation strategies seem broadly similar across languages, with grammatical "function" words dominating, the specific emphasis can shift slightly depending on the demands of the second language. In English, maintaining grammatical sentence structure becomes even more critical during hesitation, possibly leading speakers to rely even more heavily on pronouns and structural elements rather than linking devices, such as conjunctions.

Upon closer examination of the data, it became clear again that content words, such as nouns and adjectives, were almost entirely missing from the one-word repeats. Just as in Czech, hesitation clustered around the grammatical framework of the sentence rather than its content, which supported the idea stated previously - that planning pressure affects structure more than meaning in spontaneous speech production.

4.4 Length of disfluencies in L1 and L2

After analysing the frequency of disfluencies and types of repeated words, I focused on the question of how long disfluencies tend to be in individual languages. In this part of the analysis, I compared the average length of disfluencies, measured in terms of the number of words, in Czech and English monologues. This part of the study addresses Research Question 3: *Are disfluencies longer in English (L2) than in Czech (L1)?*

To answer this question, I first calculated the average disfluency length for each speaker separately in their L1 and L2, including both false starts and repeats. Then, I compared the overall averages between Czech and English to see whether operating in a second language

tends to increase the length of hesitation episodes. Table 7 below lists all specific occurrences of disfluencies, categorised by the number of words.

Length (words)	Czech FS	Czech R	English FS	English R
1	69	0	67	0
2	20	123	37	134
3	11	4	20	14
4	3	18	1	20
6	*1*	*2*	*0*	*2*

Table 7. *Distribution of disfluency lengths in L1 (Czech) and L2 (English)*

On average, false starts were slightly longer in English (1.64 words) than in Czech (1.54 words). Repeats showed a similar pattern, with English repeats averaging 2.36 words and Czech repeats averaging 2.33 words. Although the differences are small, they are consistent across both types of disfluencies. This suggests that speaking in L2 may create a slightly greater planning load, resulting in hesitation stretches that are marginally longer compared to L1.

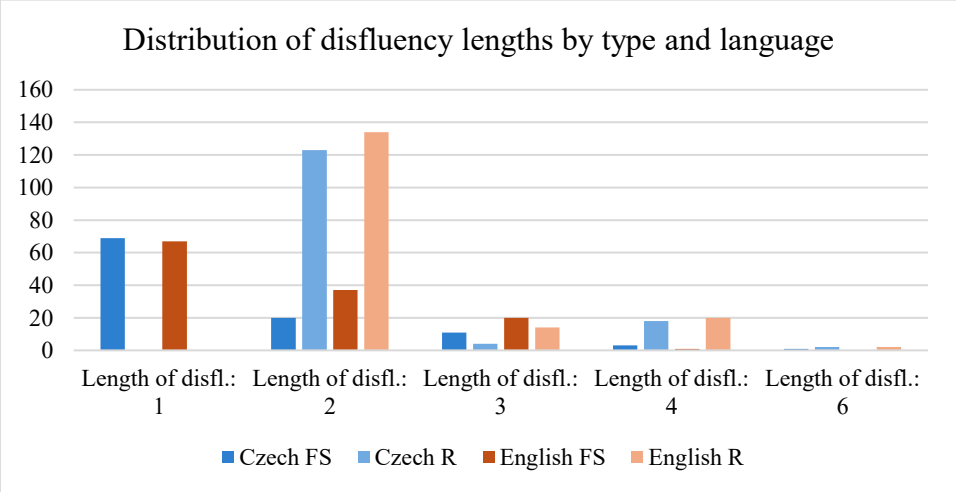

Figure 6. *Distribution of disfluency lengths by type and language*

When comparing disfluency types, repeats were notably longer than false starts in both languages. This likely reflects the structural nature of repeats: speakers often loop not just a single word, but a short phrase or group of words to buy themselves time, whereas false starts tend to interrupt speech more quickly after a smaller fragment. The fact that this difference persists across languages points to a broader pattern in speech production, where different types of disfluencies serve slightly different roles in managing planning pressure.

Overall, while the differences are not substantial, the tendency for disfluency length to increase in L2 and for repeats to involve longer segments than false starts supports the idea that hesitation

is shaped by both language proficiency and the specific strategies speakers use to stay fluent under pressure.

5 Discussion

The aim of this study was to determine whether patterns of disfluency, specifically false starts and repeats, in English as a second language reflect the disfluency patterns of Czech speakers in their first language. The results show that repeats occurred more frequently in first language speech samples (T5) than in second language samples (T1), while false starts occurred slightly more frequently in the second language. In addition, differences were found in the distribution of word classes involved in repeats, with personal pronouns being repeated more frequently in English and prepositions and conjunctions more frequently in Czech. The data did not show a clear tendency for disfluencies to be systematically longer in English than in Czech; however, structural differences between the two languages may have influenced the way speakers planned and executed repairs and repeats. These results suggest that although some transfer from L1 to L2 appears to occur, it may be influenced by structural differences between the languages and by factors related to L2 proficiency.

A comparison of these results with Gráf (2017) shows both overlaps and differences. Gráf found that repeats at the beginning of sentences, especially those of function words such as pronouns, articles, and prepositions, are common in English as a second language among Czech students and argued that such repetition may be a strategy for improving fluency rather than simply a sign of lack of fluency. In this study, the high number of repeated pronouns in English seems to be consistent with Gráf's findings. However, data from Czech as a first language in T1 showed even greater diversity of repeated elements, with prepositions and conjunctions also being abundant. This may indicate that the habits of Czech speakers in L1 are partially, but not completely, transferred to L2, likely because English syntax, with its stricter subject-verb-object word order, provides fewer opportunities to repeat prepositions and conjunctions without compromising grammatical correctness. As Gráf suggested, contact with English structures can either reinforce some patterns similar to native speech or suppress others.

A study which also corresponded with my findings was done by Peltonen (2018). She also stressed the role of individual patterns of disfluency transfer, finding that even when statistical correlations were not consistently strong, qualitative analysis revealed speaker-specific similarities between L1 and L2 stalling mechanisms. My study confirms this observation. Although quantitative differences between T1 and T5 are evident (more repeats in L1, slightly more false starts in L2), individual recordings showed considerable variation, with

some speakers repeating frequently in both languages, while others spoke relatively fluently in both languages. Due to the size of the corpus, it was not possible to study these differences between individual speakers in more detail, but they are likely to be an important factor in understanding how transfer works.

Duran-Karaoz and Tavakoli (2020) found a moderate to strong correlation between several types of disfluencies in Turkish and English, particularly in filled pauses and repairs. Although this study did not focus on filled pauses, the pattern observed in repairs (including false starts) supports the idea that some types of disfluent behaviour are consistent across languages. However, Duran-Karaoz and Tavakoli also noted that articulation rate and speech rate did not correlate between L1 and L2, suggesting that language-specific constraints influence temporal fluency more. This could explain why the distribution of repeats between T1 and T5 differs more markedly in this study: while certain types of planning problems (such as hesitation at the beginning of sentences) continue to occur, how speakers resolve these problems adapts to the possibilities of the language.

Derwing et al. (2009) found that correlations between L1 and L2 fluency were stronger in the early stages of L2 acquisition among speakers of Slavic languages than among speakers of Mandarin, and that this correlation persisted even after one year of contact with the language. They proposed that typological similarities between Slavic languages and English may enable the transfer of fluency patterns. Since Czech is a Slavic language, the results of this study are essentially compatible with their conclusions.

Lastly, de Jong et al. (2015) found a strong statistical correlation between L1 and L2 fluency measures in two language pairs (English and Turkish learners of Dutch) and argued that L1 behaviour control improves L2 fluency ratings. Their results support the idea that what appears to be fluency in L2 often has its roots in the speaker's L1 habits rather than in the speaker's level of language knowledge. The conclusions of this study are consistent with this idea, although the relatively small size of the ETC (25 speakers) and the focus on specific types of fluency rather than broader measures of fluency limit the possibility of generalising.

Several limitations of this study should be acknowledged. First, the small size of the corpus limits the strength of the conclusions, as individual differences may have a greater influence on the group results than in larger studies, given the sample size of only 25 participants. Second, the speakers in the teacher corpus were mainly from Prague and its surroundings, which means that the sample may not fully represent the language use of Czech teachers from different regions or backgrounds. Also, the recording conditions, such as the

mood, health, or stress level of the participants, were not known, which may have influenced the frequency and type of disfluencies observed. It also appears that some differences between performance in L1 and L2 can be explained by structural features of the languages themselves rather than fluency or planning ability. English syntax is less flexible than Czech and requires explicit subject-verb forms, which may partly explain why personal pronouns were so frequently repeated in T5 compared to T1.

The results suggest that L1-L2 transfer plays a role in the development of speech disfluency. Based on the data, it appears that speakers transfer some of their native disfluency strategies into English. However, given the small size of the corpus and the limited representativeness of the sample, it cannot be argued that the results provide strong or generalisable evidence of L1-L2 transfer effects. To confirm these results, further and larger studies would be necessary. However, the results of my study provide a great addition to the already existing research on language transfer and may provide a good starting point for further research on 'teacher language'.

6 Conclusion

The purpose of this research was to explore whether Czech teachers carry over habits, such as starts and repeats, from their native language to English as a foreign language. The results suggest that patterns of disfluency do not change completely when switching languages; instead, certain habits, such as the tendency to restructure utterances mid-sentence, seem to persist in both Czech and English. They also contribute to the growing body of research showing that L1 fluency habits influence L2 fluency patterns. This result is important because it supports the view that fluency in a second language is not built from scratch but is closely related to speakers' behaviour in their first language. It shows that disfluencies are not just a sign of poor language proficiency but are often a natural result of real-time planning. This finding is significant for English teachers, who may expect themselves to speak perfectly in class. Recognising that disfluencies are part of everyday speech can help shift the focus from unrealistic expectations of fluency to supporting more authentic and confident speech.

Before conducting this research, I tended to associate fluency in English with the lack of disfluencies. In my opinion, then, a "good" speaker was someone who spoke without noticeable pauses, false starts, or repeats. After examining the recordings and analysing the disfluencies, I've started to rethink this belief. When listening to the spontaneous speech of teachers who were really proficient in English, it was obvious that even very skilled speakers sometimes restart, repeat, or reshape their utterances, regardless of whether they are speaking Czech or

English. These features did not signal a lack of ability; instead, they reflected natural speech planning processes. And, to my surprise, it also worked the other way around - some of the teachers I considered to have English not suitable for someone teaching languages could have produced almost no disfluencies. As mentioned previously, some learners produce disfluencies because their "words overtake their thoughts", which, in my opinion, is not a sign of lacking fluency. If teachers were to monitor themselves when speaking English, it could lead to a slower, more careful, and less spontaneous language, which is precisely the opposite of what speakers would want.

One question remains: what now? The results of this thesis suggest that disfluency transfer from L1 is, to a certain extent, present, that speakers produce a high number of disfluencies as a coping mechanism when lost for words or when thinking, and that the main word classes that are repeated in a language are tied to the syntax of the language. Thinking of the interviews, however, I believe that the most significant implication should be that teacher education should not aim for complete smoothness but focus on managing planning difficulties naturally. Ultimately, it was not the disfluencies produced that made the teachers sound not fluent. Most teachers who produced a lot of disfluencies sounded nervous rather than inexperienced. Therefore, implications for the future should focus more on helping teachers feel confident. This could entail training in spontaneous speech and helping teachers model realistic communication. Future studies should take into account the conditions of the interviewees, their personality, as well as the corpus size. As this was a study conducted only on 25 participants, the results cannot be highly generalisable and provide solid evidence of any phenomenon.

Ultimately, fluency should not be measured by the absence of hesitation, but by the speaker's ability to continue communicating despite it. As (future) teachers, we can encourage younger speakers to feel more confident about themselves. I hope that future teacher training will place greater emphasis not only on formal accuracy but also on building confidence, spontaneity, and resilience in spoken English. This project has made me reflect on how we define fluency, how we listen to others, and how we prepare ourselves and our learners for real communication… with all its natural imperfections.

Prohlášení o použití umělé inteligence

Během vypracovávání své bakalářské práce jsem použila nástroje generativní AI na kontrolu dat a tabulek.

V Praze, dne _____

Julie Řihošková

Declaration on artificial intelligence usage

In preparation of my bachelor's thesis, I have used generative AI tools to check my data and tables.

In Prague, on the _____

Julie Řihošková

7 References

Belz, M., Sauer, S., Lüdeling, A., & Mooshammer, C. (2017). Fluently disfluent? Pauses and repairs of advanced learners and native speakers of German. *International Journal of Learner Corpus Research*, 3(2), 118-148. doi: 10.1075/ijlcr.3.2.02bel.

Clark, H. H., & Wasow, T. (1998). Repeating Words in Spontaneous Speech. *Cognitive Psychology*, *37*(3). 201-242.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition, 31*(4), 533–557. https://doi.org/10.1017/S0272263109990015.

Duran-Karaoz, Z., & Tavakoli, P. (2020). Predicting L2 fluency from L1 fluency behavior: The case of L1 Turkish and L2 English speakers. *Studies in Second Language Acquisition, 42*(4), 671–695. https://doi.org/10.1017/S0272263119000690

English Teacher Corpus. (2025). Retrieved from https://etc.ff.cuni.cz/.

Ferreira, F., Lau, E.F. and Bailey, K.G.D. (2004). Disfluencies, language comprehension, and Tree Adjoining Grammars. *Cognitive Science*, 28: 721-749. doi: 10.1207/s15516709cog2805_5.

Fox Tree, J.E. (1995). The Effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech. *Journal of Memory and Language*, *34*(6), 709-738, https://doi.org/10.1006/jmla.1995.1032.

Gráf, T. (2017). Repeats in Advanced Spoken English of Learners with Czech as L1. *Acta Universitas Carolinae - Philologica*, 3. 65-78. https://karolinum.cz/data/clanek/3906/Phil_3_2017_06_Graf.pdf.

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic Influence in Language and Cognition*. Routledge. doi:10.1558/sols.v6i3.603.

de Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language 243 fluency for first language behavior. *Applied Psycholinguistics*, *36*(2), 223–243. https://doi.org/10.1017/S0142716413000210.

Koponen, M., & Riggenbach, H. (2000). Overview: Varying perspectives on fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (p. 5–24). University of Michigan Press.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14*(1), 41–104.

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning.* Cambridge University Press. https://doi.org/10.1017/CBO9781139524537.

Peltonen, P. (2018). Exploring connections between first and second language fluency: A mixed methods approach. *The Modern Language Journal, 102*(4), 676–694. https://doi.org/10.1111/modl.12508.

Rossiter, M. J., Derwing, T. M., Manimtim, L. G., & Thomson, R. I. (2010). Oral Fluency: The Neglected Component in the Communicative Language Classroom. *Canadian Modern Language Review*, *66*(4): 583-586. doi: 10.3138/cmlr.66.4.583.

Verhoeven, L. (1989). Monitoring in bilingual speech. *Journal of Psycholinguistic Research, 18*(2), 151–163.

Williams, S. (2022). *Disfluency and proficiency in second language speech production*. Palgrave Macmillan.

Wray, A. (2002). *Formulaic Language and Lexicon.* Cambridge University Press. https://doi.org/10.1017/CBO9780511519772.

8 Resumé

1. Úvod

Tato bakalářská práce zkoumá výskyt a vlastnosti planých začátků a opakování v řeči českých učitelů angličtiny a porovnává jejich použití v mateřském jazyce učitelů (čeština, L1) a v jejich druhém jazyce (angličtina, L2). Zatímco plynulost a dysfluence jsou v rámci výzkumu prvního a druhého jazyka již dlouho studovány odděleně, vztah mezi nimi, zejména u vysoce pokročilých mluvčích, kteří nejsou rodilými mluvčími, nebyl zatím příliš prozkoumán. Na základě pozorování z osobní pedagogické praxe a příspěvků do English Teacher Corpus (ETC) si tato studie klade za cíl pochopit, zda se dysfluence jednotlivců přenášejí z jejich L1 do L2 a jak se projevují v obou jazycích. Hlavní výzkumné otázky se zabývají souvislostí mezi frekvencí dysfluencí v různých jazycích, typy slov, které se nejčastěji opakují, a potenciálními rozdíly v délce dysfluence mezi češtinou a angličtinou.

2. Teoretická část

Teoretická část nejprve vymezuje klíčové pojmy a přístupy k plynulosti a dysfluencím v řeči. Plynulost je chápána jako časový i kognitivní jev, který souvisí s rychlostí řeči, automatizací a efektivním kognitivním zpracováním. Dysfluence nejsou jen přerušeními plynulé řeči, ale plní důležité komunikační a kognitivní funkce. Tato práce se zabývá dvěma specifickými typy dysfluencí: planými začátky, kdy mluvčí přeruší větu, aby přeformuloval svou myšlenku, a opakováními, která jsou definována jako okamžitá, identická opakování slov nebo slovních spojení bez změny významu.

Dále je zkoumán jazykový transfer, konkrétně to, jak mohou rysy prvního jazyka mluvčího ovlivnit produkci druhého jazyka nad rámec gramatiky a výslovnosti a zasahovat do diskurzních strategií, jako je plynulost a dysfluence. Předchozí studie o přenosu plynulosti (např. Gráf 2017; Peltonen 2018; Derwing et al. 2009) píší o tom, že vzorce váhání, oprav a plánování řeči mohou přetrvávat napříč jazyky. Tato práce tedy vychází z názorů, že dysfluence je přirozenou součástí plánování řeči a může být ovlivněna nejen úrovní znalostí, ale také kognitivními návyky vyvinutými v prvním jazyce.

3. Materiál a metoda

Data pro analýzu pocházejí z English Teacher Corpus (ETC), sbírky nahraných rozhovorů s českými a rodilými mluvčími angličtiny, kteří vyučují angličtinu v České republice. S každým ze zúčastněných učitelů byl udělán strukturovaný rozhovor složený z pěti úkolů. Pro účely této studie byly vybrány pouze dva úkoly: spontánní monolog v angličtině (úkol 1) a

srovnatelný spontánní monolog v češtině (úkol 5). Tyto úkoly poskytly materiál vhodný pro srovnání vzorců dysfluence v L1 a L2.

Analýza se zaměřuje na nalezení a označení planých začátků a opakování v nahrávkách. Část této práce se zabývá odlišení těchto jevů od jiných typů dysfluence, jako jsou výplňová slova nebo úmyslná opakování. Aby byly dysfluence správně označeny, byly hledány jak na základě přezkoumání přepisu, tak poslechem zvukových nahrávek. Výplňková slova („uh“, „erm“, „vlastně“) nebyla zahrnuta do analýzy dysfluencí a byla zvolena pravidla pro jejich označování včetně rozhodnutí o hraničních případech.

Analyzovány byly tři hlavní charakteristiky dysfluencí:

- relativní frekvence dysfluencí na sto slov (phw),
- lexikální třída opakovaných slov (např. zájmena, slovesa, spojky)
- a průměrná délka dysfluencí měřená v slovech.

K určení statistické významnosti rozdílů mezi výsledky L1 a L2 byl použit log-likelihood test a následně t-test.

4. Výsledky

Rozbor ukázal, že učitelé obecně produkovali více dysfluencí ve svém druhém jazyce (angličtině) než v prvním jazyce (češtině). Bylo ale vypozorováno, že jednotlivci se chovali různě: mluvčí, kteří produkovali více dysfluencí v L1, měli tendenci produkovat více dysfluencí i v L2. Tento výsledek podporuje hypotézu o mezijazykovém přenosu v plynulosti řeči. Při srovnání dvou typů dysfluencí:

- plané začátky byly celkově výrazně častější než opakování.
- rozdíl mezi L1 a L2 byl výraznější u planých začátků, kdy mluvčí produkovali výrazně vyšší počet planých začátků v angličtině.
- naproti tomu v opakování byly menší rozdíly mezi oběma jazyky.

Lexikální analýza opakování ukázala, že funkční slova, hlavně zájmena, spojky a předložky, byla nejčastěji opakovanými prvky v češtině i angličtině. To naznačuje, že váhání se častěji objevuje spíše v místech gramatické struktury než ve významových slovech. Mezi oběma jazyky však byly zjištěny drobné rozdíly: anglická opakování obsahovala mírně vyšší podíl spojek a zájmen, zatímco česká opakování obsahovala více předložek, což pravděpodobně souvisí s odlišnou syntaxí obou jazyků.

Nakonec byla sledována délka dysfluencí (měřená počtem slov) a bylo zjištěno, že jejich délka byla obecně delší v angličtině. Mluvčí měli větší tendenci opakovat nebo přeformulovat delší úseky řeči, když mluvili ve svém druhém jazyce, což odráží vyšší kognitivní zátěž.5.

5. Diskuse

Výsledky poukazují na to, že strategie plánování řeči jsou napříč jazyky relativně stálé, což podporuje myšlenku, že plynulost a dysfluence řeči jsou pevně ukotveny v kognitivním stylu mluvčího. Učitelé, kteří v češtině často používali plané začátky nebo opakovali slova, měli tendenci totéž dělat i v angličtině, a to i přesto, že jejich znalost angličtiny byla na vysoké úrovni.

Rozdíl v délce a frekvenci dysfluencí mezi češtinou a angličtinou lze vysvětlit několika způsoby. Za prvé, mluvení v druhém jazyce obvykle klade vyšší kognitivní nároky, což vede k větším obtížím při plánování. Za druhé, struktura angličtiny se striktnějším slovosledem a odlišnými normami diskurzu může klást vyšší nároky na mluvčí češtiny, kteří jsou zvyklí na flexibilnější syntax jazyka. Toto potvrzují i předchozí výzkumy, které naznačují, že jazykový transfer ovlivňuje nejen gramatickou správnost, ale i faktory, jako je plynulost. Vzorce opakovaných slov odpovídají zjištěním z jiných korpusů (např. Gráf 2017), které ukazují, že funkční slova slouží jako běžná místa pro zaváhání, jelikož mluvčí plánují lexikální jednotky později než gramatické struktury během řeči v reálném čase.

Je důležité, že zatímco zvýšené dysfluence v L2 by mohly být z hlediska vnímání posluchače interpretována negativně, výsledky naznačují, že takové dysfluence nemusí nutně značit neschopnost jazyk ovládat, ale spíše odrážejí přirozené mechanismy plánování řeči.

6. Závěr

Tato studie přispívá k pochopení toho, jak se plynulost projevuje v různých jazycích, zejména u vysoce pokročilých mluvčích, kteří nejsou rodilými mluvčími, jako jsou například učitelé daného jazyka. Ukazuje, že dysfluence je ovlivněna základními kognitivními procesy, které fungují podobně v mateřském i druhém jazyce, a to i v případě pokročilé jazykové úrovně.

Výsledky mohou mít dopad na pohled na jazykové kompetence učitelů a naznačují, že občasné dysfluence, jako jsou opakování a plané začátky, by neměly být automaticky považovány za ukazatele nízké úrovně znalostí. Místo toho je lze vnímat jako běžnou a někdy i užitečnou součást spontánní řeči.

Další výzkum by mohl prozkoumat širší škálu typů dysfluencí, zaměřit se na větší vzorek mluvčích nebo analyzovat další faktory ovlivňující přenos plynulosti, jako je věk osvojení L2 nebo pedagogická praxe. Nicméně současné poznatky potvrzují, že v reálné komunikaci učitelů je plynulost složitým jevem, který je tvořen nejen jazykovou znalostí, ale také návyky z jejich mateřštiny.

9 Appendices

9.1 Instructions for Task 1

Please choose one of the following topics. You can take 2–3 minutes to think about what you are going to say, but please do not take notes. You will be expected to talk about the topic for 3-5 minutes without being given any prompts or questions. After this time, we will switch into an informal conversation.

Topic 1:

Why did you decide to become an English teacher? Are you still happy with the choice? Have you ever thought about doing something else?

Topic 2:

Have you changed as a teacher in the course of your career. How? And why?

Topic 3:

Do you remember any critical incidents in your pedagogical career which had an impact on you as a teacher? What happened and how did it affect you? What did you learn?

9.2 Instructions for Task 5

Pohovořte v 3–5 minutách v českém jazyce o tom, jak vnímáte profesi učitele. Můžete zmínit např. následující:

- co vás na práci učitele baví, a co naopak ne
- co byste ve školství rádi změnili
- zda vnímáte roli učitele jako misi či jako povolání
- jak vypadá váš běžný pracovní den
- co byste doporučili studentům učitelství s ohledem na jejich přípravu na učitelskou profesi
- atp.

9.3 Example data with marked disfluencies from Tasks 1 and 5 from speaker CZ003

9.3.1 Task 1

B: (uhu) okay alright (eh) can I start (erm) alright (erm) alright (er) so I decided to answer or to talk about the first topic (erm) (er) why (eh) and answer the question why I decided to become an English teacher (erm) because I think <R_1_2_X> it's (erm) it's rather interesting <R_1_2> that I (er) that I didn't decide to (er) become an English teacher in the first place (erm) I was very much against (er) teaching because a lot of people in my family are teachers (em) and I thought (er) that that wasn't the right path for me (erm) and when I started studying (erm) (em) (er) English at university I thought I would become a translator (erm) <FS_1> and (er) but (eh) throughout my studies I had multiple chances of (erm) teaching (er) substituting for other people (erm) and (erm) contrary to what I had planned I (er) realised that I really enjoyed the job and <FS_2> I really (erm) I was really good at it as well (er) if I may say so and so (er) I eventually came to terms with the fact that that's going to be <R_3_2> what I have (eh) what I have to do (erm) <R_1_2_P> I (er) I 'm really happy that I chose (er) to become a teacher but I'm even more happy that I chose to become an English teacher (erm) and (er) that is because of (er) (erm) multiple reasons (em) I feel like language teaching (erm) is the type of teaching that makes most the most sense to me (er) because (erm) it is very practical and skill oriented (erm) and (erm) in contrast to other colleagues of mine I can definitely tell the students (er) when they ask (em) is this gonna ever be useful (eh) in my life <R_1_2_P> I I can always say yes (erm) it is because (er) nowadays everyone (erm) is expected to speak English at a certain level (em) and so (em) I don't have to (er) let them memorise things that they're not gonna use if they choose (em) a different path (er) like it happens in other subjects right (em) also what I really like about teaching English is that (erm) <FS_1> you can (em) you have a lot of freedom in terms of what types of topics (erm) you cover as long as you cover the skills you can do that while (erm) (em) approaching a lot of different topics and I really like to: (er) get into a lot of critical thinking or current politics (erm) (eh) civil rights and things like that and (erm) I don't think I would have that freedom teaching other subjects (erm) and lastly I would say that (er) what I <R_1_2_P> really really enjoy is that (er) through speaking and through practicing speaking I get to: (em) get really close to the students and (em) I feel and <R_1_2_P> I I that has been proven right many times (eh) that I know the most about them (em) among other teachers (er) because (em) in English <R_1_2_P> we we get to share a lot and there's also this thing that (erm) in English <FS_3> students tend to (erm) very often they tend to be more honest than they would be in Czech (erm) I'm not sure why and I'm sure that <R_1_2_P> I I think there (er) someone should (erm) (em) do a study on this but when speaking English they (erm) talk much more openly about their struggles (erm) with their parents or in their relationships (er) friendships and so on (erm) and they're much more open so I really enjoy that and the last question is if I've ever (er) thought about doing something else (erm) as I said I considered being a translator (em) I also considered (er) working in H R (eh) <FS_3> it was always (er) I was always considering jobs that (er) included a lot of human interaction because that's (er) what makes me happy (er) but since I started teaching I can't imagine myself doing a different job (erm) and (erm) yeah I think I said everything is that enough

9.3.2 Task 5

B: […] myslim si že by měl učitel dostávat víc prostoru věnovat se pouze učení a méně času by měl trávit papírováním administrativou ale zároveň si myslím že třeba i dozorování na chodbách vim že v jiných zemích to je tak že dozory na chodbách nedrží učitel ale jiný nepedagogický pracovník třeba v Itálii to tak je a myslím si že by to tak mohlo být že vlastně ten učitel by potom mezi jednotlivými hodinami které odučí měl prostor si vyčistit hlavu a odpočinout si spíš než když je na chodbě a napomíná studenty ať neběhají a zároveň za ně vlastně celou dobu má zodpovědnost což bych řekla je náročná část práce <FS_2> která se o které se vlastně nemluví a která se dost podceňuje a to že v té třídě to je vlastně celkem bezpečné prostředí a učitel se věnuje tomu učení ale jak vystoupí z té třídy ať už je to prostě na chodbách na hřišti nebo při jakékoli exkurzi nebo cestě tak to pořád toto konstantní hlídání a vlastně ten pocit zodpovědnosti celou dobu <R_2_2> za zdraví (em) za zdraví těch dětí <R_1_2_V> je je <R_1_2_G> hrozně hrozně náročný já třetí bod mě trošičku pobavil protože hrozně nemam ráda <R_1_2_C> když když lidi říkají většinou v reakci na to <R_1_2_C> když když učitelé žádají lepší finanční ohodnocení nebo <FS_1> nějaký nějaké lepší uznání ve společnosti že to přece nedělají proto ale že to dělají protože to je jejich poslání tak teda moc ráda učitelství určitě moje povolání to znamená že za to očekávám nějakou finanční odměnu která bude dostatečná pro to abych vyžila ale zároveň musim říct že to vnímam trošičku i jako svoje poslání nebo misi potřebuju aby práce kterou dělám dávala nějaký smysl a myslim si že to tak má většina učitelů které znám ale nejenom učitelů myslim většina lidí kteří pracují v takových jako pomáhajících profesích nebo celkově ve službách tak jak vypadá Váš běžný pracovní den můj běžný pracovní den vypadá většinou hodně hekticky a mohla bych to popsat třeba na dnešním pracovním dnu protože dneska začal můj pracovní den už v sedum dvacet když jsem měla <FS_1> schůzi schůzku s jednou maminkou mého studenta která potřebovala probrat nějaké jeho obtíže ve škole a ta schůzka trvala do sedmi padesáti pěti takže jsem potom měla pět minut na to abych se vzpamatovala a šla učit první hodinu v tomhletom školním roce mám hodně skupin které jsou v pokročilých ročnících to znamená že končí pozdě takže skoro všechny dny končíme až osmou hodinou odpoledne a mezitím mám teda některé volné hodiny které se snažím využívat na přípravu nebo na konzultace se studenty nebo na rozhovory jako dneska ideálně i na oběd ale většinou je to tak že o přestávce moc nestíhám pořád za mnou chodí moje děti protože jsem třídní a ptají se mě <R_1_2_B|> na na nějaké věci ohledně rozvrhu <R_1_2_C> nebo nebo nějakých nadcházejících akcí nebo omluvenek nebo tak <R_1_2_C> a a <R_1_2_C> když když to učím tak většinou potřebuju třeba tři čtvrtě hodiny naprostého ticha takže třeba cestou domů ideálně vůbec nikým nemluvim nebo když potom dorazím domů tak taky ne abych si trošičku vyčistila hlavu poslední bod co bych doporučila studentům učitelství s ohledem na přípravu na učitelskou profesi tak doporučila bych jim <FS_1> aby myslim si že to nejdůležitější co bych jim doporučila je že až začnou učit a budou mít pocit že takhle to prostě nejde že se takhle nedá žít protože věnujou příliš času přípravám každá hodina je strašně vyčerpá a <R_1_2_G> strašně strašně je znervózňuje vlastně celou dobu sou v takovym pocitu nervozity <R_1_2_C> a a tak tak bych jim chtěla říct že to přejde že to opravdu přejde že mi to taky přišlo že to přece nemůže nikdy skončit <R_1_2_C> a a <FS_1> je zlepšuje se to postupně <R_1_2_C> a a mám pocit že se v té roli cítím čím dál tím líp ale řekla bych že kritický je ten první rok a <R_1_2_V> je je dobré si stanovit nějaké limity už v tom prvním roce které potom brání tomu aby člověk věnoval úplně celý svůj život té práci pro mě to třeba bylo a bylo to protože jsem měla mamku učitelku <FS_2> která u které sem některé věci jako odkoukala tak to bylo třeba ano udělám si přípravu na ten další den ale nevezmu si ty věci už domů takže já si práci domů neberu takže když sem doučila třeba v půl štvrtý tak sem třeba ještě hodinu hodinu a půl pracovala ale pak přišla chvilka kdy už sem musela jít domů a věci jsem si domů nebrala a myslim si že nastavit si <FS_2> nějaká taková nějaké takovéhle hranice hned na začátku je hodně užitečný samozřejmě se stane že prostě sou třeba i období čtvrtletek tak někdy je potřeba nebo slohovku si třeba vzít domů ale je dobré si držet nějaký prostor bez té práce trošku <FS_1> zdravé zdravý odstup <R_1_2_C> a a když se to nastaví takhle hned na začátku tak se to pak lehčeji dodržuje i v těch v těch následujících letech protože sem viděla až moc začínajících učitelů kteří prostě tady tráví čas potom do šesti a stejně si potom sbalí ten notebook a všechny ty papíry a učebnice a vezmou si to domů a dlouhodobě to není udržitelný může to chviličku fungovat ale to je úplně <FS_1> nejjasnější nejjistější cesta jak vyhořet a předpokládám že když někdo studuje učitelství tak že se připravuje na běh na dlouhou trať ne na hvězdného učitele na dva roky a pak konec takže doporučuju si stanovit hranice a meze