AUTHOR	Elina Lehtilä and Pekka Lintunen
TITLE	Individual differences in multilingual speakers' fluency: a qualitative analysis of speech and eye-movement data
YEAR	2025
DOI	10.1515/iral-2024-0195
VERSION	Accepted manuscript
CITATION	Lehtilä, Elina and Lintunen, Pekka. "Individual differences in multilingual speakers' fluency: a qualitative analysis of speech and eye-movement data" International Review of Applied Linguistics in Language Teaching, 2025. https://doi.org/10.1515/iral-2024-0195
LICENCE	CC BY-NC

Individual differences in multilingual speakers' fluency: a qualitative analysis of speech and eye-movement data

Abstract

The present study aims to examine individual differences in the connections between first (L1), second (L2), and third language (L3) utterance fluency and to investigate the potential of eye-tracking for multilingual cognitive fluency research, extending previous, mainly quantitative, research on the connections between L1 and L2/L3 fluency. Six Finnish multilingual speakers with differing L1 profiles and proficiency levels in the L2/L3 were selected from a larger dataset, and their speech and eye-movement data from picture description tasks in Finnish (L1), English (L2), and Swedish (L1/L3) were analyzed qualitatively, focusing on fluency measures and the relationship between eye movements and fluency features in the three languages. While the results indicated that some fluency measures were connected across all three languages, the results also demonstrated individual variation in the connections between L1, L2, and L3, particularly in the use of stalling mechanisms. The eye-movement data analysis showed that speakers' eye movements can provide valuable insights into the cognitive processes underlying fluency features during complex, multi-utterance descriptions in different languages. The findings have methodological implications for fluency research, indicating the need for further fluency studies involving multilingual speakers and eye-tracking, and wider implications for fluency teaching and assessment.

Keywords: cognitive fluency; eye-tracking; individual differences; multilingualism; utterance fluency

List of abbreviations:

AoA	Age of acquisition
AR	Articulation rate
AS-unit	Analysis of Speech unit
C1–4	Clusters 1–4
EVS	Eye-voice span
FP	Filled pause
FW	Filler word
L1	First language

L2	Second language
L3	Third language
LHQ3	Language History Questionnaire (version 3)
SLA	Second language acquisition
SM	Stalling mechanism
SP	Silent pause

1 Introduction

Speech fluency is a widely studied phenomenon within second language acquisition (SLA) research. While, in everyday language use, fluency is often used as a synonym for (oral) language proficiency, SLA studies have typically approached fluency from the perspective of *utterance fluency*, i.e., by analyzing the temporal features that affect the flow and continuity of speech (Segalowitz 2010; for the distinction between the definitions of fluency, see Lennon 1990). Various factors, such as the learner's first language (L1), proficiency in the second (L2) or third language (L3), and individual differences in L1 fluency have been found to contribute to L2/L3 utterance fluency (e.g., Derwing et al. 2009; Huensch & Tracy-Ventura 2017; Peltonen 2018; Peltonen & Lintunen 2022). Although a number of studies have examined the connections between L1 and L2 utterance fluency, these studies have been mainly quantitative (but see, e.g., Peltonen 2018) and mostly focused on one L1–L2 combination in a speaker's repertoire (with the exception of Peltonen & Lintunen 2022). To our knowledge, there have been no studies examining individual differences in utterance fluency across a speaker's multiple first and target languages (L2, L3, and any other subsequent languages acquired after the L1(s)). The present study approaches this under-researched area with a descriptive and in-depth analysis of six individuals' utterance fluency features in Finnish (L1), Swedish (L1/L3), and English (L2). By focusing on languages of differing typological distance (Finnish and Swedish/English) and individuals representing distinct language profiles (in terms of L1 background and L2/L3 proficiency and usage), the objective of the study is to examine the unique ways in which L1(s), L2, and L3 connect for the selected individuals (i.e., individual differences) and which factors might account for these potential individual differences.

In addition, the methods used for studying learners' *cognitive fluency* (i.e., the processes that underlie fluency-related features in speech; Segalowitz 2010) during speaking tasks have been limited to stimulated recall (Kahng 2014; Chang & Windeatt 2024). Since previous studies using eye-tracking have indicated that speakers' eye movements are informative

about the cognitive processes underlying fluency features in speech (e.g., Lee & Winke 2018; Pistono & Hartsuiker 2021), as a novel methodological approach, the present study combines eye-tracking with L1, L2, and L3 speech from the same speakers to examine the cognitive processes underlying fluency-related phenomena in picture description tasks. As there is little individual-level research on the connections in utterance fluency and between fluency features and eye movements across multilingual speakers, the study aims to provide new perspectives to the study of connections between L1, L2, and L3 fluency and to explore the potential of eye movements as an indicator of cognitive fluency in a multilingual context.

2 Literature review

2.1 Utterance fluency

Utterance fluency has typically been analyzed by measuring *speed* (e.g., articulation rate; AR), *breakdown* (frequency and duration of pauses), and *repair fluency* (e.g., frequency of corrections; e.g., Skehan 2009). Recently, positive, moderate to strong correlations have been found in both speed and breakdown fluency measures across various L1–L2/L3 combinations, but the results have not always been consistent across studies, especially regarding repair fluency (e.g., De Jong et al. 2015; Huensch & Tracy-Ventura 2017; Peltonen 2018; Duran-Karaoz & Tavakoli 2020; Peltonen & Lintunen 2022; Gao & Sun 2023; Pérez Castillejo & Urzua-Parra 2023; see also Gao & Sun 2024). Particularly strong connections have been found in end-clause silent pause (SP) measures between L1 and L2 (e.g., De Jong et al. 2015; Peltonen 2018; Gao & Sun 2023), i.e., SPs occurring between clauses or Analysis of Speech (AS) units (defined as "an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either"; Foster, Tonkyn & Wigglesworth 2000: 365). These findings have been hypothesized to reflect differences in the processes underlying SPs at different locations, with end-clause SPs generally assumed to reflect (potentially language-independent) conceptual planning, while mid-clause SPs have been associated with difficulties in message formulation and are, consequently, more likely to occur in the L2 due to limited linguistic resources (e.g., De Jong 2016a; Kahng 2014; see also Section 2.2 Cognitive fluency).

Some studies also indicate that the strength of the connections between L1 and L2 utterance fluency is, to some extent, moderated by L2 proficiency, with speakers at higher proficiency levels demonstrating more and stronger connections between L1 and L2 fluency (Huensch & Tracy-Ventura 2017; Peltonen 2018; Peltonen & Lintunen 2022; see also Suzuki & Kormos

2025). While connections in speed fluency, in particular, have been found to be robust across various L1–L2 combinations (Bradlow, Kim & Blasingame 2017), some studies suggest that cross-linguistic differences may play a role in shaping the connections between L1 and L2 utterance fluency, with potentially stronger connections between typologically related (e.g., Swedish and English in the current study) compared to unrelated languages (e.g., Finnish and English/Swedish; see Derwing et al. 2009; Huensch & Tracy-Ventura 2017; Suzuki & Kormos 2025). Therefore, the present study contributes to this line of research by exploring the same speakers' speech fluency in three languages of varying typological distance (see also Peltonen & Lintunen 2022).

Although the connections between L1 and L2 fluency have received growing attention in recent years, qualitative studies on individual differences in L1–L2 connections have been scarce. One exception comes from Peltonen (2018), who explored connections between L1 Finnish and L2 English utterance fluency across ninth graders (n = 17) and upper secondary school students (n = 25) using a mixed methods approach. In the study, the fluency features that provide the speaker time to plan their speech by means of avoiding long silences, such as filled pauses (FPs), filler words (FWs), prolongations, and repetitions, were regarded as *stalling mechanisms* (SMs) and analyzed separately from speed and breakdown fluency measures (see also Götz 2013). The qualitative examination of six speakers with the most SMs in the L2 revealed individual variation in SM use across the L1 and L2, particularly in the production of FWs and repetitions. Some speakers also showed a clear preference for one SM in their L1, but for another one in their L2, indicating idiosyncratic, potentially language-dependent patterns in the use of SMs. To date, however, there has been no qualitative research into the connections between L1, L2, and L3 utterance fluency across multilingual speakers, making the present study the first to explore this issue.

2.2 Cognitive fluency

From the perspective of the cognitive processes involved in speech production, L1 and L2 share the same general stages of conceptualizing, formulating, and articulating (Levelt 2000; Kormos 2006). *Conceptualization* involves planning and organizing the speech content, and it is considered to be largely language-independent, as opposed to *formulation* and *articulation*, where the planned message is first transformed into a language-specific form and then motorically executed. Speech production also involves a set of *self-monitoring* processes, which are used to detect potential errors in speech production and to ensure that

the speech output corresponds to the intended message. In Segalowitz's (2010; 2016) triadic framework of fluency, cognitive fluency reflects the speed and efficiency of these language-independent and language-specific processes. However, compared to utterance fluency, cognitive fluency has been somewhat under-researched, and there seems to be less consensus on operationalizing cognitive fluency (Segalowitz 2016). Typically, cognitive fluency studies involve separate tests to assess the participants' linguistic resources, such as vocabulary and grammar knowledge, and cognitive processing abilities, such as attention control and the speed of lexical access in the L2 (e.g., De Jong et al. 2013; Suzuki & Kormos 2023; 2025) and, in some cases, also in the L1 to control for individual differences in L1 cognitive fluency (Segalowitz & Freed 2004; Kahng 2020; Peltonen et al. 2024; Olkkonen, Mutta & Lintunen 2024), while measuring cognitive fluency simultaneously with speech production, as in the present study, is much rarer.

To our knowledge, stimulated recall is the only method that has been used to study the link between utterance and cognitive fluency during speech performance. For instance, Kahng (2014) compared utterance fluency in English between 15 L1 English and 31 L1 Korean L2 English speakers across different proficiency levels. A subset of the L2 English learners representing both lower (n = 9) and higher proficiency levels (n = 8) participated in stimulated recall interviews, where the participants listened to their speech performance in English and described what they were thinking of during pauses or hesitations. The results showed that the higher proficiency learners had more difficulties recalling the issues regarding L2 grammar and vocabulary than the lower proficiency learners, lending support for the hypothesis that learners at lower proficiency levels more consciously think about the issues during L2 speech production, which requires more attentional effort due to lower degree of automatization (Kormos 2006).

Recently, Chang and Windeatt (2024) also used stimulated recall to examine the causes of pauses during academic oral presentations across 22 L2 English speakers from different L1 backgrounds. The stimulated recall responses confirmed that the speakers paused both due to linguistic and cognitive problems (e.g., searching for a word or planning on what to say next), but also due to psychological factors, such as anxiety (see also Peltonen et al. 2024). Furthermore, pause location analysis corroborated the claim that some of the end-clause pauses were strategically used to, for instance, emphasize a point or provide time for the audience to process the message. However, although stimulated recall interviews can shed

light on the conscious mental processes, they are limited in that they cannot tap into the subconscious processes underlying task performance (Kahng 2020: 817), as opposed to the eye-tracking used in the present study, which provides a more direct means of investigating the more subconscious aspects of these processes as well.

2.3 Link between gaze and speech

Using visual prompts, such as comic strips, to elicit speech has a long tradition in fluency research (e.g., Derwing et al. 2009; Huensch & Tracy-Ventura 2017; Peltonen 2018; Duran-Karaoz & Tavakoli 2020; Peltonen & Lintunen 2022; Gao & Sun 2023). While eye-tracking research has shown that there is a strong connection between gaze and speech during controlled speaking tasks, such as simple action descriptions (e.g., Konopka, Meyer & Forest 2018) and object naming (e.g., Meyer et al. 2012), the connections between gaze and speech remain relatively underexplored when speakers produce complex descriptions of more natural scenes (Henderson & Ferreira 2004). Eye-tracking studies involving complex picture descriptions produced by L1 speakers of Swedish (e.g., Holšánová 2006) and English (e.g., Coco & Keller 2015) suggest that the relationship between gaze and speech becomes more complex in more realistic settings. Although gazes to objects typically precede their mention also when describing complex pictures, objects are often re-examined as they can be connected to more than one concept and grouped based on multiple principles, such as spatial or categorical proximity, similarity, and animacy (Holšánová 2006; see also Coco & Keller 2015).

In addition to the relatively limited research on eye movements during complex, multi-utterance speech, the use of eye-tracking in speech production research has mainly focused on L1 speakers, and studies on eye movements during L2 speech, particularly in less controlled speaking tasks, have been scarce (Hu & Aryadoust 2024). A notable exception comes from Lee and Winke (2018), who applied eye-tracking to examine L1 (n = 8) and L2 (n = 20) English-speaking 8–10-year-old's mental processes during a speaking test in English. Their results showed differences between the two groups in utterance fluency measures, with the L1 children demonstrating faster ARs, shorter SPs and fewer repetitions and repairs than the L2 children. The eye-movement data revealed that, particularly during SPs, FPs, and repairs, the L1 children tended to fixate on the task prompts, whereas the L2 children looked more and longer at a task-irrelevant object (a countdown timer). The authors concluded that, while the L1 children's gaze patterns seemed to be in line with previous

findings regarding the link between gaze and speech production, the L2 children's gaze patterns suggested that decreased attention to task-relevant features and, consequently, increased attention to irrelevant features may cause disfluencies in speech, particularly for young language learners. However, to our knowledge, similar research paradigms have not yet been used to examine speech fluency across adult L2/L3 speakers as in the present study.

Regarding the relationship between gaze and disfluencies in L1 speech, Griffin (2004) analyzed self-corrections produced by 33 adult L1 speakers of English in simple event or object description tasks, predicting that, compared to fluent productions, corrections would be associated with a shorter gaze time to the target object prior to naming it due to rushed word production. Contrary to the prediction, the analysis revealed that gazes to target objects before self-corrections did not differ significantly from fluent productions, but the speakers fixated on the objects significantly more after the self-corrections compared to fluent productions. Somewhat contradictory results were reported by Brown-Schmidt and Tanenhaus (2006), who conducted two experiments where 38 pairs of L1 English university students were shown arrays of shapes or objects and instructed to tell their partner to click on a highlighted item. In the trials that included a size contrast item, the speakers gazed at the contrast item significantly earlier in fluent utterances (e.g., *the small horse*) compared to disfluent ones, and the timing of the first gaze to the contrast item significantly predicted the type of disfluency, with earlier gazes associated with pre-nominal SMs (e.g., *theee uh small horse*) and later gazes with post-noun repairs (e.g., *the horse, oh the SMALL one*). Their findings provided support for the hypothesis that speakers sometimes use disfluencies strategically to buy time to reformulate the utterance (see also Clark & Fox Tree 2002). A possible explanation to the differing findings is that Griffin (2004) focused on self-corrections rather than SMs such as prolongations or FPs.

Based on Griffin (2004) and Brown-Schmidt and Tanenhaus (2006), Pistono and Hartsuiker (2021) hypothesized that examining the relationship between eye movements and disfluencies in speech, particularly in terms of the eye-voice span (EVS; the temporal distance between gaze and speech), could shed light on the processes underlying different types of disfluencies. In their study, 40 L1 Dutch university students were presented with networks of items connected by lines and asked to describe the route of a dot that traversed through the network. The authors predicted that disfluencies occurring during gazes to upcoming items would reflect stalling strategies (similarly to Brown-Schmidt & Tanenhaus

2006), while additional time gazing at an item after naming it would relate to self-monitoring processes (Griffin 2004), and disfluencies between the start of the first gaze at the picture and the onset of its name (i.e., onset-EVS) would reflect word preparation difficulties. Their results showed that all types of disfluencies were associated with longer onset-EVS, while only FPs and self-corrections predicted late gazes to items. Thus, the authors proposed that all disfluencies are associated with linguistic difficulties, but only FPs and self-corrections with self-monitoring processes. Although the studies summarized here focused on the relationship between gaze and disfluencies in L1 object naming, they provide an important point of comparison for the present study and the less controlled picture description task.

3 Current study

To summarize the gaps in fluency research outlined above, L1–L2 fluency research has mainly been limited to group-level analyses and examining fluency connections between two languages in the learner's repertoire. In addition, cognitive fluency research has almost exclusively been restricted to separate tests and the use of retrospective methods (such as stimulated recall), and eye-tracking has mostly been used to examine speech fluency across L1 speakers in relatively restricted speaking tasks. The present study aims to address these gaps by providing a detailed description of Finnish (L1), Swedish (L1/L3), and English (L2) utterance fluency across a selected group of multilingual individuals and examining the relationship between fluency features and eye movements during picture description tasks in Finnish, Swedish, and English. As both cross-linguistic differences between L1 and L2/L3 and L2/L3 proficiency have been considered to affect the strength of the connections between L1 and L2/L3 fluency, the present study aims to deepen our understanding of individual differences in the connections between L1, L2, and L3 fluency by exploring both typologically distant and similar languages (Finnish and Swedish/English) and individuals differing in terms of background characteristics related to language proficiency. To achieve these objectives, the following research questions were formulated for the present study:

RQ1. What kinds of individual differences can be identified in the connections between Finnish (L1), Swedish (L1/L3), and English (L2) utterance fluency measures across six participants representing different speaker profiles?

RQ2. How are fluency features and eye movements linked across the six participants' performances in the three languages?

To answer RQ1, the present study used multiple case study methodology to gain in-depth insights into the connections between L1, L2, and L3 utterance fluency across multilingual speakers. The method is highly recommended for the analysis of particularly complex issues, such as multilingualism, as comparisons across multiple cases allow for a more comprehensive understanding of an intricate phenomenon and, especially when combined with purposive sampling and analytic generalization, the formulation of new hypotheses and theoretical insights (e.g., Dörnyei 2007). Following Peltonen (2018), the qualitative analysis consisted of a detailed examination of the connections in Finnish, English, and Swedish utterance fluency across the six individuals differing in terms of age of acquisition, self-assessed proficiency, vocabulary knowledge, and amount of daily use of Swedish and English (see Section 3.3.1 Sampling). RQ2 was answered by conducting a qualitative sequential analysis of the semantic relations between eye movements and fluency features (see Holšánová 2006).

4 Methods

4.1 Participants

The six speakers in the present study (see Section 3.3.1. Sampling) were selected among 40 volunteers who participated in the research project "Fluency across Multilingual Speakers" (*MultiFluency*), funded by the Swedish Cultural Foundation in Finland (for the quantitative analyses of the larger dataset, see Lehtilä in preparation). All participants were multilingual, Finnish-, English-, and Swedish-speaking adults (M_{age} = 27.58, *SD* = 7.04) living in bilingual municipalities in Southern Finland. The majority of the participants were university students or graduates (*n* = 36). Most participants (*n* = 31) spoke Finnish as their L1, while 9 participants identified themselves as either simultaneous or sequential Finnish–Swedish bilinguals. All participants reported normal or corrected-to-normal vision and no language-related impairments. Before participating in the study, all participants were given a research notification and a privacy notice, containing information about the study and the processing and management of personal data. All participants gave their informed consent to participate in the study, confirming that they had received sufficient information about the study and participated voluntarily.

4.2 Materials and apparatus

The participants' linguistic background was examined using a modified version of the updated Language History Questionnaire (LHQ3; Li et al. 2020), which has been designed to assess multilingual speakers' language history and usage. The questionnaire was first translated into Finnish by the first author and then checked by the second author. The participants' vocabulary knowledge in the target languages was measured using two receptive vocabulary tests: LexTALE for English (Lemhöfer & Broersma 2012) and Sprize for Swedish (Ingves 2024). For the picture description tasks, three pictures were selected from authentic language learning and teaching materials (see Appendix 1) based on maximal comparability in terms of visual properties and the difficulty of the key vocabulary needed in the target languages (based on the Common European Framework of References levels assigned to the words in the Swedish and English KELLY lists; Kilgarriff et al. 2014). To ensure the clarity and appropriateness of the research procedure and instructions for all testing materials, the Finnish version of the LHQ3, the picture description tasks, and the vocabulary tests were piloted with multilingual adults representative of the target population prior to implementation. The piloting was deemed successful based on the participants' task performance and feedback.

The background questionnaire was administered with a secure, web-based software platform REDCap (Research Electronic Data Capture; Harris et al. 2019) hosted at the University of Turku, and the vocabulary tests were implemented with PsychoPy (Peirce et al. 2019). The eye-tracking experiment was programmed using SR Research Experiment Builder software (2.3.38). Eye movements were recorded monocularly using a desktop-mounted EyeLink 1000 eye-tracker with a sampling rate of 1000 Hz, and head position was stabilized using a chin-and-forehead rest at a viewing distance of 70 cm. Speech was recorded using a Vivitar TVM-1 telescopic video microphone connected to a Hexamix Fostex MN06 preamplifier and a low-latency, ASIO compatible Sound Blaster Audigy sound card, ensuring accurate alignment of speech and eye-movement data. The stimuli were presented on a 27" Dell Alienware AW2720HFA monitor with a refresh rate of 120 Hz and a display resolution of 1920 x 1080 pixels.

4.3 Procedure

The participants were tested individually in a quiet, windowless laboratory. Before the experiment, a 9-point calibration and a separate validation were performed for each

participant. The average error for all participants was .37° (*SD* = .09). To familiarize the participants with the apparatus and procedure, the testing began with a practice session identical to the actual experiment but with different stimuli. The participants were orally instructed that, for each task, they would first have a minute to inspect the picture and plan how they would describe it, after which they would have a maximum of two minutes to describe the picture. The instructions were first given in Finnish to ensure that all participants fully understood the task, and the first task was always done in Finnish. However, before the picture description tasks in Swedish and English, the researcher repeated the instructions in the target language to prime the language to be used and to facilitate the language switch for the participants. Before the experiment, the participants were told that they could take breaks between the tasks if needed and reminded that they could stop the experiment at any point without providing an explanation and without any negative consequences.

The experiment consisted of three picture description tasks. Each task began with a study trial, where the stimulus was presented for 60 seconds. The trial was then followed by a drift check, during which the experimenter informed the participants that they had a maximum of two minutes to describe the picture, and then the same stimulus was presented for a maximum of 120 seconds. After 120 seconds, the stimulus disappeared, but the participants were instructed that, if they had nothing else to say about the picture, they could stop the recording and continue to the following task by pressing a key. To minimize potential order and task effects, six lists were created that counterbalanced the stimuli across the participants, so that each picture appeared equally often in the Finnish, Swedish, and English task and as the first, second, and third task prompt, with the order of the target languages also counterbalanced across the participants. To maintain calibration accuracy throughout the experiment, each trial began with a drift check at the center of the screen, and a recalibration was performed if the drift check failed. After the eye-tracking experiment, the participants completed the vocabulary tests in the presence of the researcher using the laboratory computer.

4.4 Data analysis

4.4.1 Sampling

As the aim of the study was to examine speech fluency across varying speaker profiles, a maximum variation sampling procedure (Dörnyei 2007: 128) was adopted to select speakers with differing L1 profiles and proficiency levels in the target language. To identify distinct

groups of speakers with differing profiles in the data, a K-means cluster analysis was first performed in R (4.3.1; R Core Team 2023). The parameters selected for clustering were the participants' age of acquisition (AoA) in Swedish, self-assessed speaking proficiency in English and Swedish, and LexTALE and Sprize test scores, and the variables were transformed into *z* scores prior to the analysis. Different cluster solutions were considered and compared to determine the optimal number of clusters, and the four-cluster model was chosen as it delivered the most compact and well-separated clusters (based on the Dunn index; Kassambara 2017: 141). As all bilingual speakers were in Cluster 4 (C4), clusters 1–3 (C1–3) consisted of L1 Finnish speakers (see Table 1 for the summary statistics for each cluster).

[Place Table 1 near here]

Following the cluster analysis, one example case from each of the smaller clusters (1 and 2) and two example cases from each of the larger clusters (3 and 4) representing different speaker profiles were selected for the qualitative analysis. The six participants (4 female, 2 male) were between the ages of 19 and 34 years, and all except one participant were university students. The participants are referred to with gender-neutral pseudonyms to ensure their anonymity (with their cluster number and participant codes in parentheses). For Misa (C1-54), L2 English was associated with an earlier AoA and higher proficiency (in terms of self-assessed language proficiency and vocabulary test score) and amount of use compared to L3 Swedish, whereas Sumu (C2-65) can be characterized as more balanced in terms of AoA, proficiency, and amount of use in the two languages. In comparison, Viima (C3-53) and Helle (C3-62) represented more advanced L2 English speakers in terms of proficiency, with a clear difference between L2 English and L3 Swedish in both proficiency and daily amount of use. Finally, Ami (C4-71) and Ruska (C4-72) were both Finnish–Swedish bilinguals, but Ruska represented a more balanced profile across all three languages in terms of amount of use and proficiency and a higher vocabulary score in L2 English compared to Ami. The six participants' linguistic background characteristics and speech sample durations are reported in Table 2.

[Place Table 2 near here]

4.4.2 Speech and eye-movement analysis

The recordings were first automatically transcribed using Microsoft Office, after which the transcriptions were checked and corrected by a research assistant and then double-checked by the first author. The final transcripts were then automatically aligned with the audio recordings in Praat using a script (Lennes 2017), and the word boundaries were checked and adjusted. Based on previous research (e.g., Gao & Sun 2023; Peltonen & Lintunen 2022), the following measures were selected for the utterance fluency analysis: articulation rate (AR; syllables per minute of speaking time), the number and mean duration of SPs, both overall and at end-clause (i.e., at clause or AS-unit boundaries) and mid-clause positions, and the number of corrections (false starts, reformulations, and replacements; see, e.g., Peltonen et al. 2024: 11). In addition, the frequencies of the following SMs were examined: FPs (non-lexicalized pauses such as *um, öh,* or *öö* in English, Swedish, and Finnish), prolongations (one or more sounds that are markedly prolonged), FWs (discourse markers and smallwords not adding meaning to an utterance, such as *niinku, tota, tämmönen* in Finnish, *liksom, nå, nå(go)n sorts* in Swedish, or *like, well, kind of* in English), and repetitions (one or more words that are repeated without modification; see, e.g., Peltonen & Lintunen 2022). Following standard practices in utterance fluency research (e.g., Peltonen & Lintunen 2022), the fluency features were first identified aurally by the research assistant as part of the transcription process, and the final fluency feature annotations (including the identification of FWs and SP location) were conducted by the first author. The minimum SP cut-off was set at 0.25 seconds (De Jong & Bosker 2013), and all frequency measures were standardized per minute of speaking time (excluding SPs; De Jong 2016b).

For the analysis of the semantic relationship between eye movements and fluency features (RQ2), the eye-movement data were first preprocessed using the EyeLink® Data Viewer software (4.3.210) by SR Research, and the fixation report was exported in .xls format. Following initial inspection of the data, the eye-movement analysis was restricted to utterances that included specific descriptions of particular areas or objects in the picture, excluding general descriptions related to the entire picture (see, e.g., examples (1) and (2a) in Appendix 2). The speech and fixation data were then synchronized in Microsoft Excel, and each fixation that coincided with a fluency feature was coded based on its relationship with the spoken description (see Section 4.2 Results: RQ2).

5 Results

5.1 Results: RQ1

The objective of RQ1 was to examine Finnish, Swedish, and English utterance fluency for the selected six participants and interpret and contextualize the connections and differences between the three languages for each individual. In this section, the results for RQ1 are presented case-by-case, with the utterance fluency measures for the four L1 Finnish participants (C1–3) presented in Table 3, and for the two bilingual participants (C4), in Table 4.

[Place Table 3 near here]

5.1.1 Misa (C1)

Of the four L1 Finnish participants, Misa had the slowest AR and the fewest corrections in all three languages (see Table 3). When comparing the languages, L1 Finnish demonstrated the fastest AR and fewest corrections, and L3 Swedish the slowest AR and most corrections. AR also showed clear differences between all three languages, similarly to SP frequency measures, which showed that the speaker produced the fewest SPs (regardless of SP location) in L1 Finnish and the most in L3 Swedish. In contrast, all SP mean durations were the shortest in L2 English, with the most pronounced difference between L2 English and L1 Finnish/L3 Swedish in the mean duration of end-clause SPs. A closer examination of SP durations across the languages showed that end-clause SP durations, in particular, were associated with larger variation in L1 Finnish (*SD* = 0.84) and L3 Swedish (*SD* = 0.88) compared to L2 English (*SD* = 0.38). Especially the mean duration of end-clause SPs in L1 Finnish seemed to have been affected by isolated instances of exceptionally long SPs, as demonstrated in example (1) in Appendix 2.

Regarding SMs, Misa produced the fewest FPs, prolongations, and repetitions in L1 Finnish and the most in L3 Swedish, although there were relatively few occurrences of both repetitions and prolongations across the three languages. However, while the speaker used both FPs and FWs almost equally often in L1 Finnish, there were very few occurrences of FWs in L2 English and L3 Swedish, with a clear preference for FPs in both languages. A closer examination of SM location also revealed that FPs, in particular, occurred almost exclusively in connection with SPs in L2 English (95.0% of all FPs) and L3 Swedish (88.2%

of all FPs; compared to 55.6% in L1 Finnish; see examples (2a), (2b), and (2c) in Appendix 2).

5.1.2 Sumu (C2)

Sumu also had the fastest AR in L1 Finnish, but, compared to Misa (C1), ARs were very similar in L2 English and L3 Swedish. Sumu was one of the speakers in the sample with the most and, on average, longest SPs, with the fewest and shortest SPs overall in L1 Finnish and the most and longest in L3 Swedish. However, in L2 English, Sumu produced the most end-clause SPs and the fewest mid-clause SPs, and, likewise, the fewest corrections, while the frequencies of end-clause SPs and corrections were more similar between L1 Finnish and L3 Swedish. Regarding SMs, in L1 Finnish, Sumu used mostly FWs (particularly the word *niinku/niinkun*), FPs to some extent, and very few prolongations. In L2 English, FWs (particularly the word *like*, the English equivalent of *niinku*) and prolongations were the most frequent SMs, and FPs were also frequently used. In comparison, FWs were very infrequent in L3 Swedish, while FPs and prolongations were used fairly evenly. Repetitions were used rarely, and only in the target languages. Sumu's use of different SMs across the three languages is demonstrated in examples (3a), (3b), and (3c) in Appendix 2.

5.1.3 Viima (C3)

Viima was one of the fastest speakers in the sample in all three languages, and showed, again, the fastest AR in L1 Finnish, but, similarly to Sumu (C2), very similar ARs in the target languages. Viima was also one of the participants with the shortest SPs on average, with relatively small differences in the mean SP duration measures across the three languages. However, in terms of SP frequency, Viima produced the most SPs in total in L1 Finnish and the fewest in L2 English, and SP location analysis showed that this difference was mostly due to the high frequency of mid-clause SPs in L1 Finnish (and, to a lesser extent, in L3 Swedish) compared to L2 English. Viima also produced the most corrections in all three languages: the highest number of corrections was found in L3 Swedish, while L1 Finnish and L2 English were very similar in terms of correction frequency. In addition, while corrections in L1 Finnish and L2 English were typically isolated false starts, replacements, or reformulations, the corrections in L3 Swedish tended to cluster together, as demonstrated in examples (4a) and (4b) in Appendix 2.

Regarding SMs, FWs were the most frequent SM in L1 Finnish (particularly the word *totanoinni*; a dialectal variation of the filler *tota*, with no direct equivalent in English), closely followed by FPs, which were the most frequent SM in both L2 English and L3 Swedish. The frequency of FWs was the lowest in L2 English, while repetitions were more frequent in the target languages compared to the L1. The frequency of prolongations was also the highest in L3 Swedish, but, altogether, the speaker produced very few prolongations across the three languages. Examples (5a), (5b), and (5c) in Appendix 2 demonstrate Viima's use of different SMs across the three languages.

5.1.4 Helle (C3)

Of the six participants, Helle had the fastest AR in L1 Finnish, while, similarly to Sumu (C2) and Viima (C3), the ARs in the target languages were very similar. However, the frequency of SPs was the highest in L3 Swedish, mostly due to the high frequency of mid-clause SPs. In addition, and, again, similarly to Viima, the frequency of mid-clause SPs was the lowest in L2 English. There were relatively small differences between the mean durations of all SPs across the languages, but, on average, end-clause SPs were longer in L1 Finnish compared to the target languages, while mid-clause SPs were longer in L3 Swedish compared to L1 Finnish and L2 English. Although Helle produced more corrections in L1 Finnish and L3 Swedish compared to L2 English, a closer examination of repair context showed that, in L1 Finnish and L2 English, all corrections were relatively short reformulations, replacements, or false starts, whereas, in L3 Swedish, corrections also spanned longer stretches of speech (see examples (6a), (6b), and (6c) in Appendix 2). As also exemplified in (6a) and (6c), the only SM Helle used in L1 Finnish were FWs, which were also used to a similar extent in L3 Swedish. In L2 English, Helle used all types of SMs more evenly, with prolongations being the most frequent SM (see (6b)). Prolongations were slightly less frequent in L3 Swedish, whereas FPs were the most frequent SM. Similarly to Viima, repetitions were also relatively frequent in both target languages, although these were mostly very short repetitions of the conjunction *and* (*och* in Swedish), as demonstrated in examples (7a) and (7b) in Appendix 2.

5.1.5 Ami (C4)

[Place Table 4 near here]

In terms of AR, Ami spoke the fastest in L1 Finnish, but, in contrast to the L1 Finnish participants, AR was the slowest in L2 English, with a clear difference between L1 Finnish and L1 Swedish/L2 English (see Table 4). The frequency of all SPs was the lowest in L1 Finnish and the highest in L2 English, although their mean durations were slightly shorter in L1 Finnish and L2 English compared to L1 Swedish. SP location analysis showed that L1 Swedish was very close to L1 Finnish with regards to the frequency of end-clause SPs, but closer to L2 English in mid-clause SP frequency. In contrast, mid-clause SPs were longer in L1 Swedish compared to L1 Finnish and L2 English, with larger differences between the languages than in the mean durations of end-clause SPs, which were very similar across all three languages. The frequency of corrections was the lowest in L1 Swedish and the highest in L2 English, but the majority of corrections in L1 Finnish and L2 English were relatively short false starts. In contrast, and, similarly to Helle (C3), corrections in L1 Swedish included lengthier reformulations that required more backtracking (see examples (8a), (8b), and (8c) in Appendix 2). The examples also illustrate Ami's use of SMs: in the L1s, FWs were the most frequent SM, although clearly more dominant in L1 Finnish than in L1 Swedish. While repetition frequencies were very similar in the L1s, FPs were more frequent in L1 Swedish and L2 English compared to L1 Finnish. Prolongations were also more frequent in L2 English compared to L1 Finnish, although relatively rarely used in L1 Swedish as well.

5.1.6 Ruska (C4)

For Ruska, AR was also the fastest in L1 Finnish, but, in contrast to all other participants, AR was very similar in L1 Swedish, with a clearly slower AR in L2 English compared to the L1s. Of all the participants, Ruska produced the fewest corrections in all three languages, with very minor differences between the languages. The frequency of all SPs was the highest in L1 Finnish and the lowest in L2 English. Their mean duration was also the longest in L1 Finnish, but the shortest in L1 Swedish. However, SP location analysis showed that the difference in frequency was due to a difference in the number of mid-clause SPs, which was the highest in L1 Finnish but very similar between L1 Swedish and L2 English. In contrast, the difference in mean SP duration seemed to be more related to the production of end-clause SPs, with L1 Finnish demonstrating, on average, longer end-clause SPs compared to L1 Swedish and L2 English. However, the mean duration of end-clause SPs in L1 Finnish was associated with larger variation (*SD* = 0.90) compared to L1 Swedish (*SD* = 0.46) and L2 English (*SD* = 0.64), again, similarly to Misa (C1), owing to a few very long end-clause SPs in L1 Finnish,

as demonstrated in example (9a) in Appendix 2, with more typical end-clause SP durations shown in example (9b).

Regarding SMs, Ruska only used FWs, FPs, and prolongations in the three languages, but all three types of SMs were relatively evenly used in L1 Finnish, whereas a clearer preference was found for prolongations over FPs, and, to a lesser extent, FWs in L1 Swedish. In comparison, prolongations were also the most frequent SM in L2 English, but FPs were more frequent than FWs. Ruska's use of different SMs is illustrated in examples (10a), (10b), and (10c) in Appendix 2.

To summarize the results of RQ1, connections between the three languages in AR seemed to be less prone to individual variation than connections in repair frequency, SP measures, and SMs. In addition, the bilingual participants displayed stronger connections between Finnish and Swedish than the L1 Finnish participants in terms of AR, although this connection was more pronounced for Ruska (potentially due to more similar daily amount of use and self-assessed proficiency in the L1s). Repair frequency also showed connections between the three languages, particularly for the speakers with the fewest (Misa and Ruska) and most (Viima) corrections. SP duration measures also seemed to demonstrate connections between the three languages, although particularly the more advanced L2 English speakers (in terms of self-assessed proficiency and vocabulary test score) produced both fewer and shorter mid-clause SPs in L2 English than in their L1(s). While individual differences in the connections between L1, L2, and L3 could be observed for all fluency measures, particularly the connections in SM use seemed to be largely idiosyncratic and unrelated to the participants' linguistic background factors. The potential factors influencing the differences in fluency connections for the six individuals are discussed further in the Discussion.

5.2 Results: RQ2

The qualitative analysis of the relationship between speech and eye movements revealed that, during picture description tasks, the speakers' gaze was often ahead of speech. When fluency features occurred in the speech flow (in the form of SPs, corrections, or SMs), the participants were typically examining the areas they were either about to describe next or currently describing. For all participants, these gazes were associated with all types of fluency features (end-clause and mid-clause SPs, corrections, and SMs) in all three languages. Two examples of these gaze patterns are shown in Figure 1, with examples (6a)

and (7b) from Helle (C3). In both examples, the speaker's gaze was on the area they were describing during all fluency features, but, in the L1 Finnish example, Helle's gaze shifted to one of the adults in the scene during the mid-clause SP preceding the reformulation *lapset ja aikuiset leikkii* ('children and adults are playing'), whereas, in the L2 English example, Helle's gaze was on the adult in the pool already during the first SM (FP *er*).

[Place Figure 1 near here]

On less frequent occasions, the speakers' gaze still lingered on the objects they had already finished describing. For most participants, these gazes were more frequent during fluency features in the target languages compared to the L1(s), with only Helle showing a somewhat opposite pattern (with few instances in both L1 Finnish and L2 English, but none in L3 Swedish). The differences between the L1 and the target languages were the most pronounced for Misa (C1) and Viima (C3): for Misa, these "late" gazes were very rare in L1 Finnish (identified in only 3.3% of all fluency features) and L2 English (3.2%) compared to L3 Swedish (14.3%), whereas, for Viima, these gazes were more frequent in both target languages (L2 English: 9.3%, L3 Swedish: 9.5%) compared to L1 Finnish (0.0%). These gazes were almost exclusively associated with end-clause SPs and SMs, as demonstrated in Figure 2 with the example (2b) from Misa. The gaze data showed that Misa's gaze was on the described area during the SMs, SPs, and the replaced *mannen* ('the man') that occurred during the description. Then the gaze lingered on the described object during the following end-clause SP and FP, although Misa did not continue describing the object.

[Place Figure 2 near here]

All participants also demonstrated gaze patterns in all three languages where the object of their gaze seemed to have no direct link to their current description. For both bilingual participants (C4), these gazes were less frequent in L1 Swedish (Ami: 5.6%, Ruska: 14.5%) compared to L1 Finnish (Ami: 21.9%, Ruska: 31.7%) and L2 English (Ami: 21.1%, Ruska: 29.3%). A closer examination of these gazes revealed that they were often "anticipatory" in a sense that the participants were examining objects they were going to describe later, as demonstrated in panel A of Figure 3, with example (9b) from Ruska. The gaze data showed that, during the description of the swimmers, Ruska's gaze was first, for the most part, on the children on the beach. Then, during the first end-clause SP, the gaze shifted to the couple

walking on the beach, which Ruska began describing after finishing the description of the children on the beach. In panel B of Figure 3, the example (8a) from Ami shows that their gaze was primarily on the objects they were currently describing (the swimmers), and also occasionally on the objects to be described next (the dolphins), but, during the mid-clause SP prior to the very short false start, Ami also briefly fixated on the seagulls, which they began describing after finishing the description of the dolphins.

[Place Figure 3 near here]

In contrast to the bilingual speakers, for most of the L1 Finnish participants, the gazes that had no apparent link to their current description were more frequent during fluency features in the target languages compared to their L1. However, for Misa, the proportions were very similar across the target languages (L2 English: 24.2%, L3 Swedish: 23.2%), whereas, for Helle, these gazes were slightly more frequent in L3 Swedish (7.1%) compared to L1 Finnish (4.4%) and L2 English (5.0%). In contrast, for Viima, these gazes were more frequent in L1 Finnish (16.7%) and L3 Swedish (14.3%) compared to L2 English (5.6%). For Sumu (C2), the differences were clearer across all three languages, with a higher proportion of these gazes in L2 English (32.1%) compared to L3 Swedish (16.9%) and L1 Finnish (10.8%). In addition to the "anticipatory" gazes, these gazes could also be on previously described objects that were not returned to again during the task, as demonstrated in panel A of Figure 4 with example (4b) from Viima. During the first false start, Viima's gaze was on the area they were currently describing, but, during the following false starts, the gaze moved to areas that they had already described (the pool, the locker room, and the infirmary), but did not return to describing during the task. Finally, these gazes could also be associated with areas that the speakers never described during the task. On the right side of Figure 4, example (3b) from Sumu shows that their gaze was, during most of the fluency features, on areas other than the one they were currently describing (the swimmers). These included both objects that Sumu had described previously (the children and the couple on the beach) and that they would describe later (the seagulls), but also objects that they never described during the task (the dolphins).

[Place Figure 4 near here]

6 Discussion

The qualitative analysis focusing on individual differences in the connections between L1 (Finnish/Swedish), L2 (English) and L3 (Swedish) utterance fluency measures (RQ1) revealed more individual variation in L1–L2/L3 connections in breakdown fluency and SMs than in speed and repair fluency. In line with previous studies (Huensch & Tracy-Ventura 2017; Peltonen 2018; Peltonen & Lintunen 2022; Pérez Castillejo & Urzua-Parra 2023), there were clear connections between AR in the L1 and the target languages for the four L1 Finnish speakers, with slower/faster ARs in the L1 associated with a similar AR profile in the target languages. In addition to the balanced L2/L3 speaker Sumu, the ARs of the two advanced L2 English speakers were very similar in both target languages, indicating that they were able to maintain a similar AR in the target languages regardless of the proficiency distance between the L2 and L3, although potentially at the expense of using more mid-clause SPs (for Helle) or corrections (for Viima). In comparison, the two bilingual participants demonstrated a slower AR in L2 English compared to their L1s, indicating that bilingual speakers may exhibit connections in speed fluency between their L1s regardless of the typological distance between languages (see also Bradlow, Kim & Blasingame 2017). However, the connection between the L1s was more pronounced for Ruska, potentially due to the more balanced amount of use and self-assessed proficiency in the L1s compared to Ami.

The results also indicated that repair frequency is, to some extent, connected between L1, L2, and L3, with the two speakers who made the fewest corrections in their L1(s) also making the fewest corrections in the L2 and L3 (Misa, C1 and Ruska, C4), despite differences in L2/L3 proficiency. Of all participants, correction frequency was also the highest in all three languages for Viima (C3), although much higher in the L3 compared to the L1 and L2. However, compared to speed fluency, connections in repair fluency across the three languages seemed to be more prone to individual variation (see also Peltonen et al. 2024). For Sumu (C2) and Helle (C3), correction frequency was lower in L2 English compared to L1 Finnish and L3 Swedish, and, for Ami (C4), higher in the L2 compared to the L1s. For Sumu, however, the frequency of end-clause SPs was much higher in the L2 compared to the L1 and L3, suggesting that they may have been engaged in more conceptual planning during the task in the L2, potentially leading to a decreased need for repairing speech. In comparison, for both Helle and Ami, the repairs made in L1 Finnish and L2 English were more similar in terms of extent, while repairs made in L3/L1 Swedish included longer utterances that were rejected or reformulated. These individual differences in repair fluency across L1, L2, and L3

may partially explain why connections between L1 and L2 correction frequency have been found in some studies (De Jong et al. 2015; Duran-Karaoz & Tavakoli 2020), but not in others (Huensch & Tracy-Ventura 2017; Gao & Sun 2023; Pérez Castillejo & Urzua-Parra 2023; Peltonen et al. 2024; Suzuki & Kormos 2025). The findings also suggest that, in addition to frequency measures, repair fluency analysis might benefit from a more detailed examination of repair extent, which has been somewhat neglected in past fluency studies (see also Bosker et al. 2013: 172).

For breakdown fluency measures, the findings demonstrated somewhat mixed patterns for all six participants. In line with previous studies (De Jong et al. 2015; Peltonen 2018; Gao & Sun 2023), SP duration measures seemed to indicate connections between L1, L2, and L3. However, for some participants (Misa, Helle, and Ruska), end-clause SPs were notably longer in Finnish compared to Swedish and/or English. For Misa and Ruska, end-clause SP mean duration in L1 Finnish seemed to have been affected by isolated, markedly longer SPs, likely reflecting conceptualization processes and potentially arguing in favor of setting an upper limit for SP duration used in some previous studies (e.g., Peltonen 2018; Peltonen & Lintunen 2022). In addition, particularly Misa used FPs much more frequently in the target languages (possibly reflecting an overuse of FPs in the target languages; see Götz 2013) and almost exclusively in connection with SPs in L2 English, potentially helping to avoid long SPs (see, e.g., Peltonen 2018).

Furthermore, for all participants, mid-clause SPs were more frequent and/or longer in L1 Finnish/L1 Swedish compared to L2 English. These findings are in contrast to previous L2 fluency studies, which have found that mid-clause SPs tend to be longer and more frequent in L2 speech (e.g., Kahng 2014; De Jong 2016a). As previous research has indicated that speakers tend to engage in more extensive speech planning in the L2 than in the L1 (Konopka, Meyer & Forest 2018), a potential explanation for the findings is that the speakers used the planning time more efficiently in the L2 compared to the L1(s), thus leading to fewer and/or shorter mid-clause SPs during the speaking task in the L2. However, this hypothesis should be explored in future studies by comparing the use of planning time and its potential effects on utterance fluency across speaking tasks in the L1(s) and target languages.

The results regarding SMs indicated language-dependent idiosyncratic patterns in the use of SMs across the participants, consistent with previous findings on individual variation in SM

use in both L1 and L2 speech (Götz 2013; Kahng 2014; Peltonen 2018). For instance, a clear preference for FWs in L1 Finnish was found for Sumu, Helle, and Ami (also in L1 Swedish, to a lesser extent), but the SMs used in the target languages were more varied, and none of the speakers displayed a similar overuse of a single SM in the target languages. In contrast, Misa, Viima and Ruska displayed similar, more varied SM profiles in their L1(s), but seemed to prefer a specific SM in the target languages (FPs for Misa and Viima, prolongations for Ruska). These findings are in line with Peltonen's (2018) study, and may partially explain the mixed results regarding connections in SMs between L1 and L2/L3, particularly in terms of FPs and repetitions (e.g., De Jong et al. 2015; Peltonen & Lintunen 2022; Gao & Sun 2023; Pérez Castillejo & Urzua-Parra 2023). In addition, for all L1 Finnish participants, repetition frequencies were higher and more similar in the target languages compared to the L1, potentially indicating that certain SMs may show stronger connections between target languages (regardless of language proficiency) rather than connections between the L1 and L2/L3 (see also Peltonen & Lintunen 2022).

Regarding the associations between fluency features and eye movements during picture description in the three languages (RQ2), the speakers' gaze during SPs, corrections, and SMs was revealing in terms of whether the participants were examining areas they were currently describing or going to describe next, presumably formulating their description, or examining areas that had no direct link with their current or upcoming description. The gazes to areas the participants were currently describing or about to describe were associated with all types of fluency features across all three languages, which could be interpreted as lending support for Pistono and Hartsuiker's (2021) hypothesis that, also during L2/L3 speech, different kinds of fluency features can be associated with linguistic encoding difficulties related to formulation processes. Examples from the gaze data also provided support for the hypothesis that fluency features are sometimes used strategically to incorporate newly detected information, and earlier gazes to relevant referents seemed to be associated with SMs (such as FPs and repetitions) and later gazes with repairs (Brown-Schmidt & Tanenhaus 2006). However, as unambiguous examples of these gaze patterns were relatively rare due to the unstructured nature of the task, this hypothesis should be examined further in connection with different task types involving more linear narratives, such as comic strip descriptions (e.g., Derwing et al. 2009; Peltonen & Lintunen 2022).

In comparison, gazes that were not directly connected with the participants' current or upcoming description suggested that sometimes the participants were processing areas they had described earlier or that they would describe later, more likely reflecting processes related to self-monitoring and conceptualization rather than formulation. In addition, as illustrated in the examples, these gazes also co-occurred with corrections in the data, suggesting that corrections may also occur due to decreased attention to relevant areas, implying rushed message preparation (cf. Griffin 2004). Particularly in the target languages, sometimes these gazes were also on objects the speakers never described during the task, suggesting that they were distracted by irrelevant and/or potentially problematic objects, possibly aware of their importance regarding the completion of the task but unable to find the vocabulary to describe them in the target language (see also Kahng 2014; Lee & Winke 2018). Although less frequent in the data, gazes to objects the speakers had just finished describing were also more common during the target languages for most participants and almost exclusively associated with end-clause SPs and SMs, potentially lending support for the hypothesis that some fluency features are more likely to be associated with self-monitoring processes (Pistono & Hartsuiker 2021). However, it should be noted that, while some participants displayed much higher proportions of gazes to areas that were not related to their current or upcoming description in the target languages (Misa and Sumu), for some participants the proportions were very similar in L1 Finnish and L2 English (Viima, Ami, and Ruska) or across all three languages (Helle). These findings might suggest individual variation in the extent to which speakers engage in conceptualizing or self-monitoring both in the L1(s) and L2/L3, indicating the need to further examine individual differences in cognitive fluency in both L1(s) and target languages across multilingual speakers (see also Segalowitz & Freed 2004; Kahng 2020; Peltonen et al. 2024; Olkkonen, Mutta & Lintunen 2024).

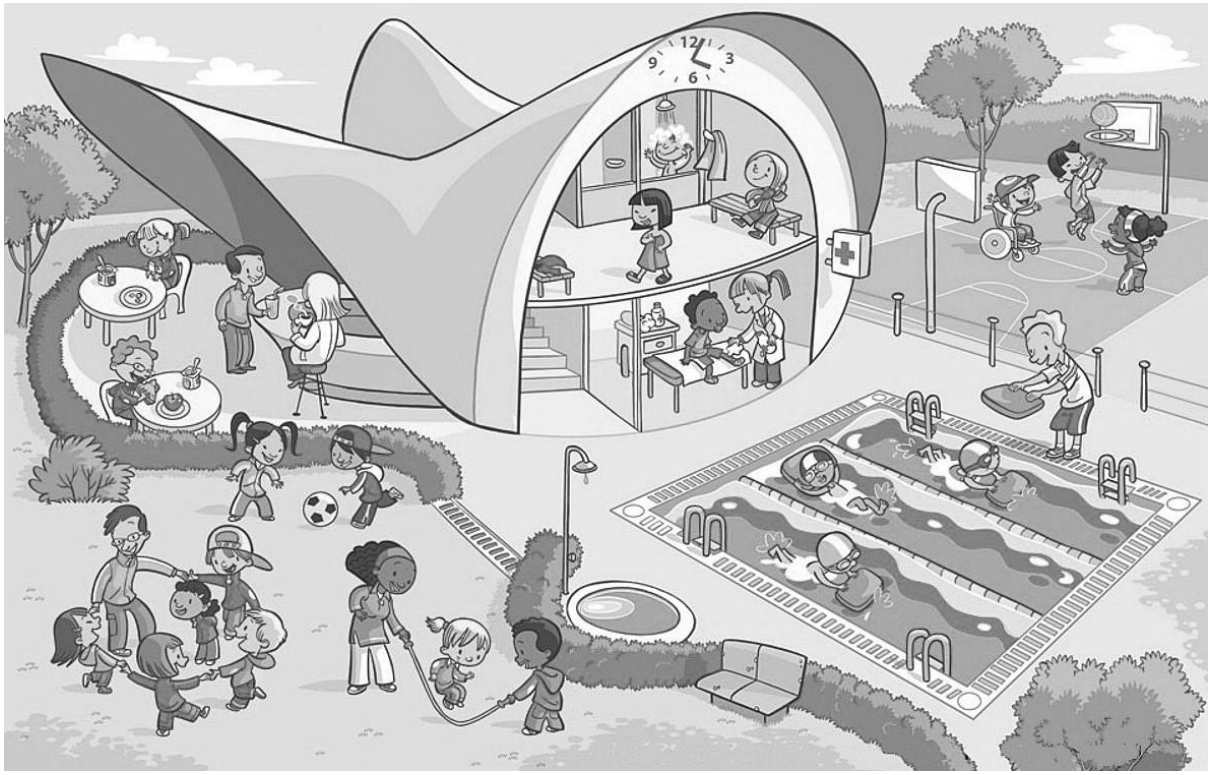
7 Conclusion

To our knowledge, the present study has been the first to include both speech from the same speakers in three languages and eye-tracking to examine individual differences in the connections in utterance fluency and the cognitive processes underlying fluency features during speech production. The results showed that there is considerable individual variation in the connections between L1, L2, and L3 utterance fluency, and, particularly in the use of SMs, the connections were seemingly idiosyncratic and not mediated by linguistic background factors, such as the speaker's L1(s) or target language proficiency. However, due

to the general limitations related to small sample sizes in qualitative research, our findings may not be generalizable or representative of a larger population. Future studies should, therefore, explore individual differences in fluency connections between L1(s) and target languages, especially regarding different SM profiles, with larger sample sizes and different multilingual populations. In addition, in our study, the speakers' language proficiency was only measured in terms of self-assessment and vocabulary knowledge in the target languages. The role of proficiency as a factor affecting individual differences in the connections between L1(s) and target languages should be examined more systematically in future. While eye-movement data proved to be useful for drawing inferences about the cognitive processes underlying fluency features during picture description tasks, the method is naturally limited to speaking tasks involving visual prompts. The findings also showed that the relationship between gaze and speech becomes more complex and open to interpretation in less controlled speaking tasks (see also Holšánová 2006; Coco & Keller 2015), indicating that future fluency studies could benefit from the triangulation of speech and eye-movement data with retrospective verbal reports, such as stimulated recall.

Despite the limitations, the qualitative approach adopted in the present study was particularly valuable in providing an in-depth, holistic understanding of the individual variation associated with the connections between L1, L2, and L3 fluency. The findings also raise important methodological considerations for future utterance fluency research, particularly regarding the interconnected nature of fluency measures, which potentially contributes to explaining the somewhat inconsistent findings regarding the connections between L1 and L2/L3 fluency in previous quantitative studies. Another contribution is the novel use of eye-tracking to examine the same speakers' cognitive processes during picture descriptions in different languages, which can provide the initial methodological basis for future multilingual fluency and SLA studies combining eye-tracking with spontaneous speech production. The results also demonstrate the importance of addressing individual variation in fluency in both L2 teaching and assessment contexts: for instance, as all participants used FWs more in their L1(s) than in the L2 and L3, raising learners' awareness of the SMs they use in their L1(s) and teaching FWs appropriate in the target language might help learners to use varying types of SMs more efficiently and to avoid overreliance on a single SM. The results also highlight the need to extend fluency analyses to multilingual speakers and beyond group-level analyses to gain a deeper understanding of the factors affecting fluency across individuals in both L1(s) and target languages.

Appendix 1. The picture prompts used in the picture description tasks, illustrated by Carmén Martin Ortega and reproduced with permission from the publisher Editorial Everest.



Appendix 2. Examples from the speech data with the relevant fluency features highlighted in bold and the location of silent pauses marked with the abbreviations E (end-clause) and M (mid-clause). For examples in Finnish and Swedish, an English translation (translated by the first author) is provided in italics below the original example.

Misa

(1) varmaa (0.63, M) m:eteli on melkoinen **(3.97, E)** voi olla myös että popkorni tuoksuu
*probably (0.63, M) the n:oise is quite loud **(3.97, E)** it might also be that popcorn smells*

(2a) **(1.39, E) uhm** (.) people are having quite a fun time it seems **(0.33, E) uh (0.88, E)**

(2b) **(1.25, M) öö** (.) med (0.48, M) hjälp (0.66, M) av (0.45, M) m:annen- öm mannet där **(1.18, E) öö (2.04, E)**
(1.25, M) uh** (.) with (0.48, M) help (0.66, M) of (0.45, M) the m:an- um man there **(1.18, E) uh (2.04, E)

(2c) aika **ömm** mielenkiintonen tämä arkkitehtuuri (1.04, E) siihen on (1.42, M) käytetty paljon rahaa
*quite **um** interesting this architecture (1.04, E) a lot of money has been (1.42, M) spent on it*

Sumu

(3a) kaikil on **niinku** jotain tekemistä (1.08, E) kukaan ei oo **niinku** (1.22, M) vaan yksin
*everyone has **like** something to do (1.08, E) no one is **like** (1.22, M) just alone*

(3b) (0.76, E) and then (1.64, M) there's **like** people swimmi**ng:** (1.93, E) **uh**

(3c) se**n:** (2.88, M) **öö** (0.85, M) **sen** finns d**e:**t (2.77, M) **mm** (0.32, M) barn so**m**: (0.88, M) simmar
*the**n:** (2.88, M) **uh** (0.85, M) **then** th**e:**re are (2.77, M) **mm** (0.32, M) kids wh**o**: (0.88, M) are swimming*

Viima

(4a) uh **there's a paren-** (0.80, M) there's a family going in (.) and (0.26, E)

(4b) ja **där en-** (.) **är det- so-** som jag sade där är (0.61, M) ynga (0.78, M) barn som (.) spelar (0.27, M) basketboll
*yeah **there is-** (.) **is there- a-** as I said there are (0.61, M) young (0.78, M) kids who (.) are playing (0.27, M) basketball*

(5a) **totanoinni tämmönen** (1.15, M) vanhus katselee **totanoinni** (.) televi- televisiota (0.42, M) **öö** yksinään
***FW** this **kind of** (1.15, M) old man is watching **FW** (.) televi- television (0.42, M) **uh** alone*

(5b) on the left (.) we have seen **we have (.) seen** that **uh** there's a cinema

(5c) sen är- (0.48, M) klockan är nästa**n nästan öö** fem- (0.37, M) **aa** halv **halv** fem (0.35, E) **aa** (0.58, E)
*then it's- (0.48, M) it's almost **almost uh** five- (0.37, M) **uh** half **half** past four (0.35, E) **uh** (0.58, E)*

Helle

(6a) siellä on **tommonen** (0.63, M) nurmikenttä missä jotku **lapset leikkii-** (0.76, M) lapset ja aikuiset leikkii
*there's a **kind of** (0.63, M) grass field where some **children are playing-** (0.76, M) children and adults are playing*

(6b) **okay** we're looking at **a: s:**tay at the **seashore-** (0.28, M) seaside **kind of thing**

(6c) **och den där (0.68, M) äm (1.54, M) man som har (0.58, M) öö (1.02, M) klär av sig (0.52, M) som-** (0.84, M) **öm** (1.37, M) **ja okej** jag kommer inte ihåg vad heter den

and that (0.68, M) um (1.54, M) man who has (0.58, M) uh (1.02, M) undresses (0.52, M) as- *(0.84, M)* ***um*** *(1.37, M)* ***yeah okay*** *I can't remember what that's called*

(7a) plats att köpa popcorn **o:**ch (.) **och** ett bibliotek och en konstgaller**i:**
place to buy popcorn ***a:****nd (.)* ***and*** *a library and an art galler****y:***

(7b) with (.) **er** (.) kids and (0.36, M) **and** adults swimming in the swimming pool

Ami

(8a) siinä on **tota** (0.69, M) meressä (0.64, M) muutama **e-** (0.57, M) tyyppi uimassa
there ***FW*** *(0.69, M) in the sea (0.64, M) there's a few* ***a-*** *(0.57, M) guys swimming*

(8b) **and se-** and then **uhm** (1.51, M) on the **uh** (1.66, M) other side there are children who are playing (0.27, M) on the field

(8c) det finn**s:** (0.39, M) **aa** (.) barn som **aam** (1.16, M) **spelar där-** (1.41, M) **aa** vad heter det (1.70) **aa nå** kastar bollen där
*there a****re:*** *(0.39, M)* ***uh*** *(.) children who* ***um*** *(1.16, M)* ***are playing there-*** *(0.)* ***uh*** *what is it called (1.70, M)* ***uh well*** *are throwing the ball there*

Ruska

(9a) kaikki näyttää aika ilosilta **(4.53, E)** näyttää lämpimältä **(2.52, E)**
everyone looks quite happy ***(4.53, E)*** *it looks warm* ***(2.53, E)***

(9b) ihmisii jotka ui meressä **(0.51, E)** lapsii jotka tekee jotain hiekka (1.15, M) rakennelmia rannalla **(0.70, E)** vanhempi pariskunta kävelemässä (.)
people swimming in the sea ***(0.51, E)*** *kids doing some sand (1.15, M) constructions on the beach* ***(0.70, E)*** *an older couple walking (.)*

(10a) **mm** tossa **tota** (.) ton (0.46, M) **niinku** (0.37, M) terassialueen (0.50, M) aidalla istuu lokki

***mm** there **FW** (.) on that (0.46, M) **like** (0.37, M) terrace area (0.50, M) fence a seagull is sitting*

(10b) och sen nere (0.67, M) **nån sorts** (0.25, M) läkare**s:** (1.57, M) rum
*and then downstairs (0.67, M) **some sort of** (0.25, M) doctor**s:** (1.57, M) room*

(10c) **mm** (1.33, E) to the right there seems to be **a:** (.) art museum **of some sort**

8 References

Bosker, Hans Rutger, Anne-France Pinget, Hugo Quené, Ted Sanders & Nivja H De Jong. 2013. What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing* 30(2). 159–175. https://doi.org/10.1177/0265532212455394.

Bradlow, Ann R., Midam Kim & Michael Blasingame. 2017. Language-independent talker-specificity in first-language and second-language speech production by bilingual talkers: L1 speaking rate predicts L2 speaking rate. *The Journal of the Acoustical Society of America* 141(2). 886–899. https://doi.org/10.1121/1.4976044.

Brown-Schmidt, S & M Tanenhaus. 2006. Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language* 54(4). 592–609. https://doi.org/10.1016/j.jml.2005.12.008.

Chang, Heejin & Scott Windeatt. 2024. Fluency issues in L2 academic presentations: Linguistic, cognitive and psychological influences on pausing behaviour. *Language Teaching Research* 1–24. https://doi.org/10.1177/13621688241233718.

Clark, Herbert H. & Jean E. Fox Tree. 2002. Using uh and um in spontaneous speaking. *Cognition* 84(1). 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3.

Coco, Moreno I. & Frank Keller. 2015. Integrating mechanisms of visual guidance in naturalistic language production. *Cognitive Processing* 16(2). 131–150. https://doi.org/10.1007/s10339-014-0642-0.

De Jong, Nivja H. 2016a. Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching* 54(2). 113–132. https://doi.org/10.1515/iral-2016-9993.

De Jong, Nivja H. 2016b. Fluency in second language assessment. In Dina Tsagari & Jayanti Banerjee (eds.), *Handbook of Second Language Assessment*, 203–218. Boston/Berlin: De Gruyter. https://doi.org/10.1515/9781614513827-015.

De Jong, Nivja H. & Hans Rutger Bosker. 2013. Choosing a threshold for silent pauses to measure second language fluency. In Robert Eklund (ed.), *The 6th Workshop on Disfluency in Spontaneous Speech (DiSS 2013)* (TMH-QPSR), vol. 54, 17–20. Stockholm, Sweden. http://www.isca-speech.org/archive/diss_2013/papers/dis6_017.pdf.

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

De Jong, Nivja H., Margarita P. Steinel, Arjen Florijn, Rob Schoonen & Jan H. Hulstijn. 2013. Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics* 34(5). 893–916. https://doi.org/10.1017/S0142716412000069.

Derwing, Tracey M., Murray J. Munro, Ronald I. Thomson & Marian J. Rossiter. 2009. The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition* 31(04). 533–557. https://doi.org/10.1017/S0272263109990015.

Dörnyei, Zoltán. 2007. *Research methods in applied linguistics: Quantitative, qualitative and mixed methodologies*. Oxford: Oxford University Press.

Duran-Karaoz, Zeynep & Parvaneh Tavakoli. 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/S0272263119000755.

Foster, Pauline, Alan Tonkyn & Gillian Wigglesworth. 2000. Measuring spoken language: A unit for all reasons. *Applied Linguistics* 21(3). 354–375. https://doi.org/10.1093/applin/21.3.354.

Gao, Jianmin & Peijian Paul Sun. 2023. To correct or not: The role of L1 fluency in understanding and measuring L2 fluency. *TESOL Quarterly* 57(2). 643–655. https://doi.org/10.1002/tesq.3198.
Gao, Jianmin & Peijian Paul Sun. 2024. How does learners' L2 utterance fluency relate to their L1? A meta-analysis. *International Journal of Applied Linguistics* 34(1). 276–291. https://doi.org/10.1111/ijal.12493.
Götz, Sandra. 2013. *Fluency in native and nonnative English speech* (Studies in Corpus Linguistics). Vol. 53. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/scl.53.
Griffin, Zenzi M. 2004. The eyes are right when the mouth is wrong. *Psychological Science* 15(12). 814–821. https://doi.org/10.1111/j.0956-7976.2004.00761.x.
Harris, Paul A., Robert Taylor, Brenda L. Minor, Veida Elliott, Michelle Fernandez, Lindsay O'Neal, Laura McLeod, et al. 2019. The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics* 95. 103208. https://doi.org/10.1016/j.jbi.2019.103208.
Henderson, John M. & Fernanda Ferreira. 2004. Scene perception for psycholinguists. In John M. Henderson & Fernanda Ferreira (eds.), *The interface of language, vision, and action: Eye movements and the visual world*, 1–58. New York, NY, US: Psychology Press.
Holšánová, Jana. 2006. Dynamics of picture viewing and picture description. In Liliana Albertazzi (ed.), *Advances in consciousness research*, vol. 67, 235–256. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/aicr.67.17hol.
Hu, Xin & Vahid Aryadoust. 2024. A systematic review of eye-tracking technology in second language research. *Languages* 9(4). 141. https://doi.org/10.3390/languages9040141.
Huensch, Amanda & Nicole Tracy-Ventura. 2017. Understanding second language fluency behavior: The effects of individual differences in first language fluency, cross-linguistic differences, and proficiency over time. *Applied Psycholinguistics* 38(4). 755–785. https://doi.org/10.1017/S0142716416000424.
Ingves, Anna. 2024. *Vägar mot ett svenskt ordförråd : Nyanlända ungdomars ordförrådsutveckling på språkintroduktionsprogrammet*. Uppsala: Uppsala University PhD Thesis.
Kahng, Jimin. 2014. Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning* 64(4). 809–854. https://doi.org/10.1111/lang.12084.
Kahng, Jimin. 2020. Explaining second language utterance fluency: Contribution of cognitive fluency and first language utterance fluency. *Applied Psycholinguistics* 41(2). 457–480. https://doi.org/10.1017/S0142716420000065.
Kassambara, Alboukadel. 2017. *Practical guide to cluster analysis in R: Unsupervised machine learning*. London: STHDA.
Kilgarriff, Adam, Frieda Charalabopoulou, Maria Gavrilidou, Janne Bondi Johannessen, Saussan Khalil, Sofie Johansson Kokkinakis, Robert Lew, Serge Sharoff, Ravikiran Vadlapudi & Elena Volodina. 2014. Corpus-based vocabulary lists for language learners for nine languages. *Language Resources and Evaluation* 48(1). 121–163. https://doi.org/10.1007/s10579-013-9251-2.
Konopka, Agnieszka E., Antje Meyer & Tess A. Forest. 2018. Planning to speak in L1 and L2. *Cognitive Psychology* 102. 72–104. https://doi.org/10.1016/j.cogpsych.2017.12.003.
Kormos, Judit. 2006. *Speech production and second language acquisition*. 1st edn. New York: Routledge. https://doi.org/10.4324/9780203763964.

Lee, Shinhye & Paula Winke. 2018. Young learners' response processes when taking computerized tasks for speaking assessment. *Language Testing* 35(2). 239–269. https://doi.org/10.1177/0265532217704009.
Lehtilä, Elina. in preparation. Exploring connections between L1, L2, and L3 fluency using speech and gaze data: A quantitative perspective.
Lemhöfer, Kristin & Mirjam Broersma. 2012. Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods* 44(2). 325–343. https://doi.org/10.3758/s13428-011-0146-0.
Lennes, Mietta. 2017. SpeCT - Speech Corpus Toolkit for Praat (v1.0.0). First release on GitHub. Zenodo. https://doi.org/10.5281/ZENODO.375923.
Lennon, Paul. 1990. Investigating fluency in EFL: A quantitative approach. *Language Learning* 40(3). 387–417. https://doi.org/10.1111/j.1467-1770.1990.tb00669.x.
Levelt, Willem J. M. 2000. Producing spoken language: A blueprint of the speaker. In Colin M. Brown & Peter Hagoort (eds.), *The Neurocognition of Language*, 82–122. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198507932.003.0004.
Li, Ping, Fan Zhang, Anya Yu & Xiaowei Zhao. 2020. Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition* 23(5). 938–944. https://doi.org/10.1017/S1366728918001153.
Meyer, Antje S., Linda Wheeldon, Femke Van Der Meulen & Agnieszka Konopka. 2012. Effects of speech rate and practice on the allocation of visual attention in multiple object naming. *Frontiers in Psychology* 3. 1–13. https://doi.org/10.3389/fpsyg.2012.00039.
Olkkonen, Sanna, Maarit Mutta & Pekka Lintunen. 2024. Fluency profiles in L2 monologue production: the role of underlying cognitive factors. In Minnaleena Toivola, Pekka Lintunen & Leena Maria Heikkola (eds.), *New directions in speech research – Freely produced speech as data.* (AFinLA-Teema), vol. 17, 167–190. Jyväskylä: The Finnish Association for Applied Linguistics (AFinLA). https://doi.org/10.30660/afinla.136716.
Peirce, Jonathan, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman & Jonas Kristoffer Lindeløv. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51(1). 195–203. https://doi.org/10.3758/s13428-018-01193-y.
Peltonen, Pauliina. 2018. Exploring connections between first and second language fluency: A mixed methods approach. *The Modern Language Journal* 102(4). 676–692. https://doi.org/10.1111/modl.12516.
Peltonen, Pauliina & Pekka Lintunen. 2022. Multilingual speakers' L1, L2, and L3 fluency across languages: A study of Finnish, Swedish, and English. *Nordand* 17(1). 48–63. https://doi.org/10.18261/nordand.17.1.4.
Peltonen, Pauliina, Sanna Olkkonen, Magdalena Szyszka & Pekka Lintunen. 2024. L2 repair fluency through the lenses of L1 repair fluency, cognitive fluency, and language anxiety. *Applied Linguistics Review* 1–23. https://doi.org/10.1515/applirev-2023-0011.
Pérez Castillejo, Susana & Katherine Urzua-Parra. 2023. Predicting L2-Spanish fluency from L1-English fluency and L2 proficiency: A conceptual replication. *Studies in Second Language Acquisition* 45(4). 1090–1103. https://doi.org/10.1017/S0272263122000535.

Pistono, Aurélie & Robert J. Hartsuiker. 2021. Eye-movements can help disentangle mechanisms underlying disfluency. *Language, Cognition and Neuroscience* 36(8). 1038–1055. https://doi.org/10.1080/23273798.2021.1905166.
R Core Team. 2023. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
Segalowitz, Norman. 2010. *Cognitive bases of second language fluency*. New York: Routledge.
Segalowitz, Norman. 2016. Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching* 54(2). 79–95. https://doi.org/10.1515/iral-2016-9991.
Segalowitz, Norman & Barbara F. Freed. 2004. Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition* 26(2). 173–199. https://doi.org/10.1017/S0272263104262027.
Skehan, P. 2009. Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics* 30(4). 510–532. https://doi.org/10.1093/applin/amp047.
Suzuki, Shungo & Judit Kormos. 2023. The multidimensionality of second language oral fluency: Interfacing cognitive fluency and utterance fluency. *Studies in Second Language Acquisition* 45(1). 38–64. https://doi.org/10.1017/S0272263121000899.
Suzuki, Shungo & Judit Kormos. 2025. The moderating role of L2 proficiency in the predictive power of L1 fluency on L2 utterance fluency. *Language Testing* 42(1). 73–99. https://doi.org/10.1177/02655322241241851.

Table 1. The means and standard deviations of the variables used in the clustering analysis for each cluster (C1–4).

Variable	**C1: Higher L2; lower L3 (*n* = 9)**	**C2: Balanced L2 & L3 (*n* = 6)**	**C3: Advanced L2; lower L3 (*n* = 13)**	**C4: Bilinguals; high L2 (*n* = 12)**
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Age of acquisition in Swedish	13.00 (0.00)	12.33 (1.21)	11.38 (2.06)	0.75 (1.86)
Speaking proficiency in Swedish (1–7)[a]	3.44 (0.88)	4.17 (1.60)	3.46 (0.78)	5.83 (0.94)
Speaking proficiency in English (1–7)[a]	5.33 (0.71)	4.17 (0.75)	6.23 (0.60)	5.83 (0.83)
Vocabulary test score in Swedish (0–100%)	56.44 (8.92)	78.33 (12.93)	68.15 (11.57)	94.33 (9.36)
Vocabulary test score in English (0–100%)	76.94 (6.03)	68.75 (8.87)	96.25 (3.19)	86.04 (7.72)

[a]Based on the participant's self-rated language proficiency on a scale between 1 = "very poor" – 7 = "excellent" (Li et al. 2020)

Table 2. The six speakers' (clusters 1–4; C1–4) background characteristics and speech recording durations in Finnish (FI), English (EN), and Swedish (SW).

Background variable	**Misa (C1)**			**Sumu (C2)**			**Viima (C3)**			**Helle (C3)**			**Ami (C4)**			**Ruska (C4)**		
	L1 FI	L2 EN	L3 SW	L1 FI	L2 EN	L3 SW	L1 FI	L2 EN	L3 SW	L1 FI	L2 EN	L3 SW	L1 FI	L1 SW	L2 EN	L1 FI	L1 SW	L2 EN
Age of acquisition (years)	0	9	13	0	9	10	0	8	11	0	7	11	0	0	4	0	0	9
Daily amount of use (%)	80	15	5	50	25	25	60	25	5	50	50	0	50	40	10	33	33	33
Language proficiency (1–7)[a]	7.00	5.75	4.75	6.00	4.75	4.50	6.50	5.75	3.50	7.00	5.75	3.50	7.00	6.00	6.00	6.50	6.75	6.50
Vocabulary test score (0–100%)	-	78.75	60.00	-	70.00	75.00	-	91.25	64.00	-	95.00	82.00	-	-	75.00	-	-	92.50
Speech sample duration (s)	120.0	120.0	120.0	120.0	120.0	120.0	94.3	120.0	107.2	55.6	64.7	60.0	119.0	101.8	110.8	120.0	120.0	120.0

[a]Based on the participant's self-rated language proficiency on different components of a language (listening, speaking, reading, and writing on a scale between 1 = "very poor" – 7 = "excellent" (Li et al. 2020)

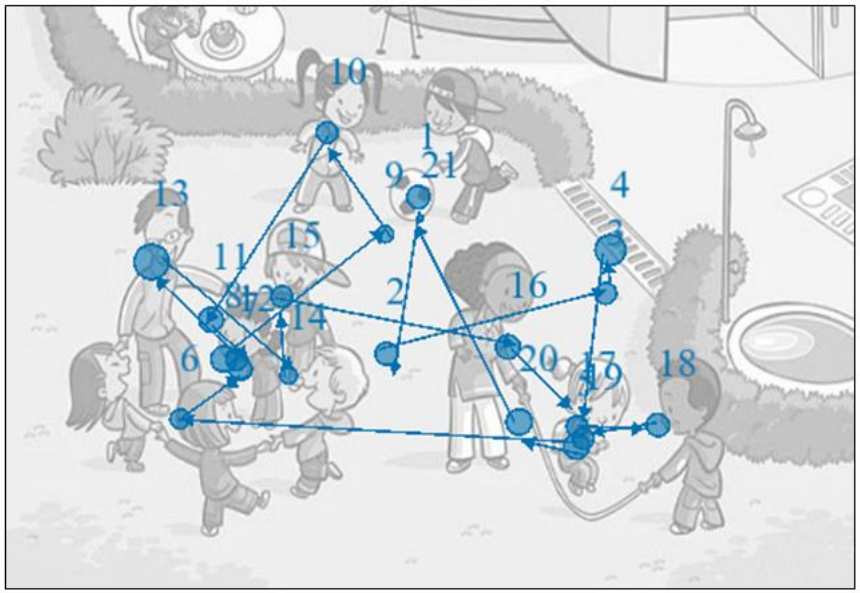
Table 3. Measures for the four L1 Finnish speakers (clusters 1–3; C1–3) in Finnish (FI), English (EN), and Swedish (SW).

Measure	**Misa (C1)**			**Sumu (C2)**			**Viima (C3)**			**Helle (C3)**		
	L1 FI	L2 EN	L3 SW	L1 FI	L2 EN	L3 SW	L1 FI	L2 EN	L3 SW	L1 FI	L2 EN	L3 SW
Articulation rate	305.00	223.00	157.50	346.09	232.61	228.58	364.59	301.97	292.47	401.14	264.93	260.40
Silent pauses (SPs) / min. of speaking time (ST)	24.75	42.00	59.47	44.37	53.44	63.43	39.97	31.74	38.15	27.15	28.18	43.40
Mean duration of SPs (s)	1.09	0.75	1.18	1.05	1.22	1.32	0.57	0.62	0.63	0.66	0.61	0.73
End-clause SPs / min. of ST	11.65	22.91	30.84	17.75	36.67	25.13	16.54	17.16	19.07	6.17	12.68	12.40
Mean duration of end-clause SPs (s)	1.29	0.81	1.53	1.22	1.32	1.54	0.72	0.67	0.68	1.02	0.80	0.77
Mid-clause SPs / min. of ST	13.10	19.09	28.64	26.62	16.76	38.30	23.43	14.58	19.07	20.98	15.50	31.00
Mean duration of mid-clause SPs (s)	0.92	0.68	0.81	0.93	1.01	1.17	0.46	0.55	0.58	0.55	0.45	0.72
Corrections / min. of ST	2.18	2.29	3.30	8.87	4.19	7.18	10.34	9.44	16.69	6.17	4.23	7.75
Filled pauses / min. of ST	6.55	15.27	18.72	7.10	10.48	10.77	11.72	14.58	15.10	0.00	4.23	10.85
Prolongations / min. of ST	1.46	1.53	3.30	3.55	12.57	9.57	1.38	1.72	2.38	0.00	7.05	6.20
Filler words / min. of ST	5.10	0.76	2.20	20.41	13.62	2.39	13.78	4.29	6.36	9.87	5.64	9.30
Repetitions / min. of ST	0.00	0.76	2.20	0.00	2.10	2.39	1.38	4.29	5.56	0.00	5.64	6.20

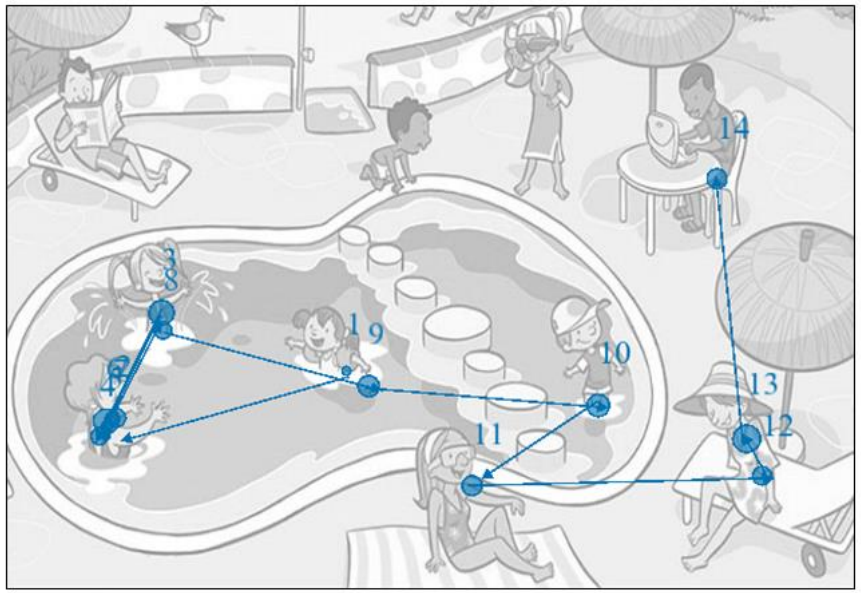
Table 4. Measures for the two bilingual speakers (cluster 4; C4) in Finnish (FI), Swedish (SW), and English (EN).

Measure	**Ami**			**Ruska**		
	L1 FI	L1 SW	L2 EN	L1 FI	L1 SW	L2 EN
Articulation rate	316.46	240.05	210.00	307.10	298.07	261.70
Silent pauses (SPs) / min. of speaking time (ST)	31.90	35.75	43.58	35.62	31.53	29.24
Mean duration of SPs (s)	0.77	0.87	0.74	0.98	0.82	0.86
End-clause SPs / min. of ST	12.93	11.92	19.81	19.79	22.93	20.68
Mean duration of end-clause SPs (s)	0.86	0.84	0.84	1.15	0.85	0.97
Mid-clause SPs / min. of ST	18.97	23.83	23.77	15.83	8.60	8.56
Mean duration of mid-clause SPs (s)	0.71	0.89	0.65	0.77	0.74	0.60
Corrections / min. of ST	5.17	4.26	7.13	1.58	0.72	1.43
Filled pauses / min. of ST	1.72	8.51	8.72	5.54	2.87	3.57
Prolongations / min. of ST	0.86	2.55	5.55	5.54	8.60	7.13
Filler words / min. of ST	25.01	13.62	3.96	7.92	5.73	1.43
Repetitions / min. of ST	4.31	3.40	1.58	0.00	0.00	0.00

A



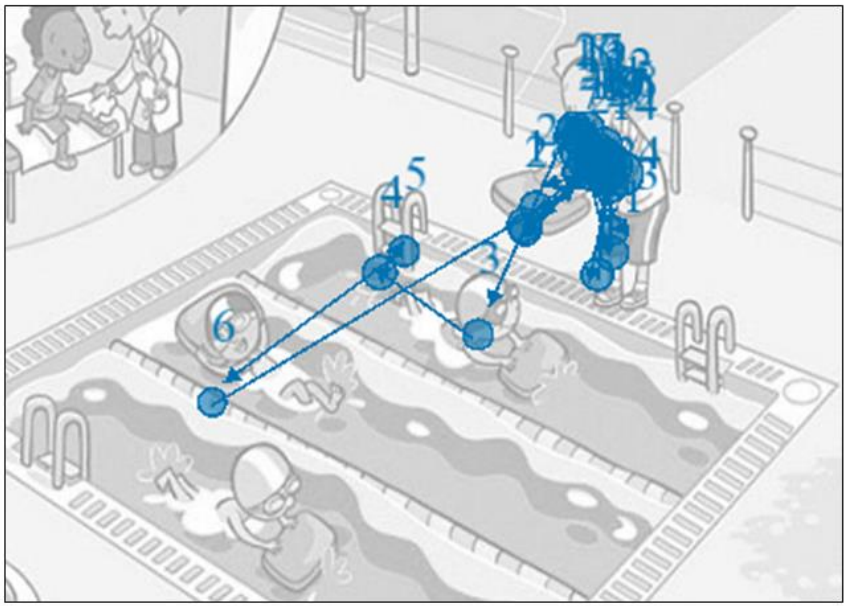
B



Transcription	siellä	on	**tommonen**	**(0.63, M)**	nurmikenttä	missä	jotku				
Fixation_index	1	2	**3**	**4**	**5**	6	7	8	9	10	
	lapset	**leikkii-**	**(0.76, M)**	lapset	ja	aikuiset	leikkii				
	11	**12**	**13**	**14**	**15**	16	17	18	19	20	

with (.)	**er**	(.)	kids	and	**(0.36, M)**	**and**	adults	swimming	in the		
1	**2**	**3**	4	5	6	7	**8**	**9**	10	11	12
swimming	pool										
13	14										

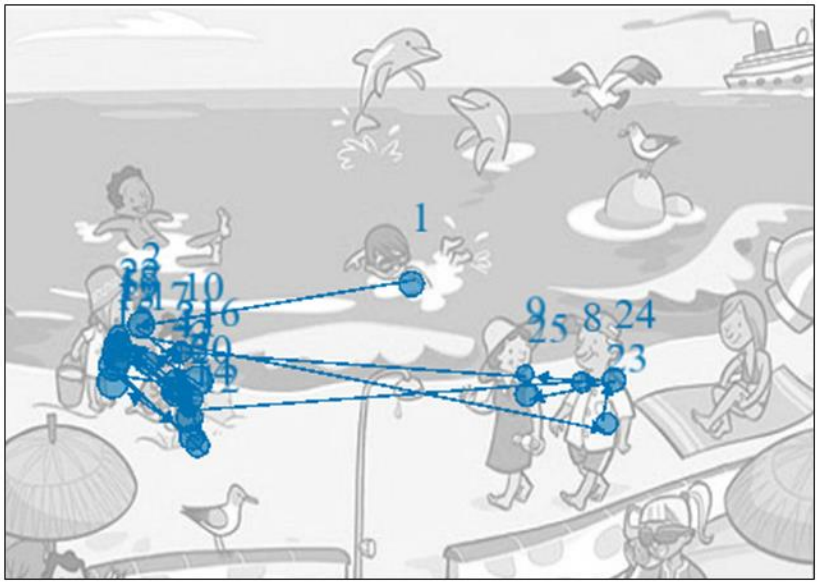
Figure 1. Gaze data from Helle (C3), with the example (6a) shown in panel A (English translation: *there's a **kind of (0.63, M)** grass field where some **children are playing- (0.76, M)** children and adults are playing*) and the example (7b) in panel B, generated using the EyeLink® Data Viewer software by SR Research. Fixation_index numbers below the transcriptions refer to the fixation numbers in the images (pictures illustrated by Carmén Martin Ortega and reproduced with permission from the publisher Editorial Everest).



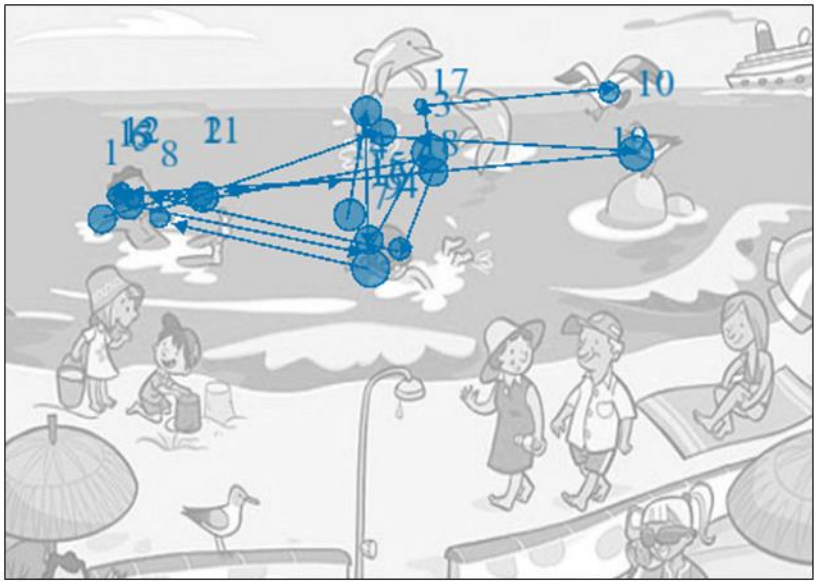
Transcription	**öö**	(.)	med	**(0.48, M)**	hjälp	**(0.66, M)**	av	**(0.45, M)**	**m:annen**		
Fixation_index	**1**	2	3	**4**	**5**	6	7	**8**	9	**10**	**11–18**
	öm		mannet	där	**(1.18, E)**	**öö**					
	19		20	21	**22–26**	**27**					

Figure 2. Gaze data with the example (2b) from Misa (C1; English translation: ***uh*** *(.) with (0.48, M) help (0.66, M) of (0.45, M) the m:an- um man there* ***(1.18, E) uh***), generated using the EyeLink® Data Viewer software by SR Research. Fixation_index numbers below the transcription refer to the fixation numbers in the image (picture illustrated by Carmén Martin Ortega and reproduced with permission from the publisher Editorial Everest).

A



B



Transcription	ihmisii	jotka	ui	meressä	**(0.51, E)**	lapsii	jotka	tekee	jotain
Fixation_index	1 2	3	4	5	6	**7–9**	10	11	12	13
	hiekka	**(1.15, M)**	rakennelmia	rannalla					
	14	**15–19**	20	21	22	23–25			

	siinä	on	**tota**	**(0.69, M)**	meressä	**(0.64, M)**	muutama	**e-**	**(0.57, M)**
	1	2	**3**	**4**	**5**	**6**	**7**	8	**9**	**10**	**11**	12	13	**14**	**15**	**16**
	tyyppi	uimassa														
	17	18–21														

Figure 3. Gaze data with the example (9b) from Ruska (C4) shown in panel A (English translation: *people swimming in the sea **(0.51, E)** kids doing some sand **(1.15, M)** constructions on the beach*) and the example (8a) from Ami (C4) shown in panel B (English translation: *there **FW** (0.69, M) in the sea (0.64, M) there's a few **a-** (0.57, M) guys swimming*), generated using the EyeLink® Data Viewer software by SR Research. Fixation_index numbers below the transcriptions refer to the fixation numbers in the images (picture illustrated by Carmén Martin Ortega and reproduced with permission from the publisher Editorial Everest).

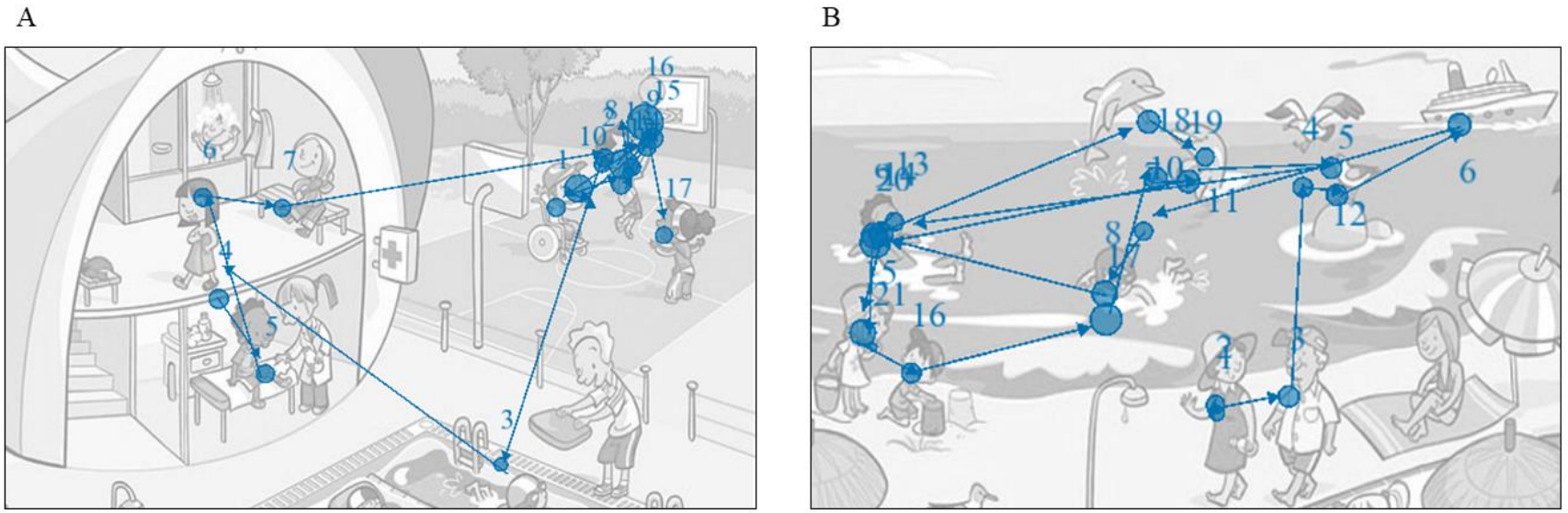


Figure 4. Gaze data with the example (4b) from Viima (C3) shown in panel A (English translation: ***yeah there is-*** *(.)* ***is there- a-*** *as I said there are* ***(0.61, M)*** *young* ***(0.78, M)*** *kids who (.) are playing* ***(0.27, M)*** *basketball*) and the example (3b) from Sumu (C2) shown in panel B, generated using the EyeLink® Data Viewer software by SR Research. Fixation_index numbers below the transcriptions refer to the fixation numbers in the images (pictures illustrated by Carmén Martin Ortega and reproduced with permission from the publisher Editorial Everest).