Fluency assessment

Incorporating syntactic distance in a new measure of pause location

Sylvain Coulange and Nivja H. de Jong
Université Grenoble Alpes | Leiden University

Speech fluency is often assessed using articulation rate and pause frequency. However, not all pauses hinder fluency: when placed strategically, they structure discourse and enhance comprehensibility. To better characterize speaker fluency, it is crucial to consider *where* pauses occur. Traditional approaches rely on categorical syntactic boundaries (e.g., clauses or phrases), but inadequately capture syntactic complexity. We propose a continuous measure of pause placement based on syntactic distance between adjacent words. Using spontaneous English speech from Japanese learners and native speakers, we show that syntactic distance robustly predicts both pause location and duration across proficiency levels. We compare its contribution to proficiency classification against baseline and categorical models. The syntactic distance model outperforms all others, explaining 87% of variance (versus 65% for baseline and 76% for clause/phrase models), with strongest model fit and lowest prediction error. This measure provides a robust and meaningful predictor of L2 speech fluency.

Keywords: fluency, syntactic complexity, pauses, syntax, L2 speech, assessment

1. Introduction

Consider the following utterance: "While I was jogging this morning // I saw a tiny // dog playing with a ball." While both pauses (represented with "//") in this example may have identical duration, their impact on perceived fluency differs considerably. The pause between "morning" and "I" occurs at a natural clause boundary, supporting discourse structure, while the pause between "tiny" and "dog" interrupts the flow within a noun phrase and, besides specific emphatic reasons, can potentially lead to misunderstanding. Hence, depending on where

https://doi.org/10.1075/jslp.25042.cou | Published online: 13 March 2026 | Corrected: 30 March 2026
Journal of Second Language Pronunciation ISSN 2215-1931 | E-ISSN 2215-194X

they occur, pauses can either contribute to structuring speech or conversely have a detrimental effect on comprehension.

Speech fluency is commonly conceptualized as encompassing breakdown (pauses and hesitations), speed (articulation rate), and repair (false starts and repetitions, Tavakoli & Skehan, 2005). In second language (L2) contexts, fluency is typically measured using speech rate, pause frequency, and mean pause duration (Shea & Leonard, 2019). However, these global measures fail to capture the nuanced relationship between pause placement and syntactic structure. As Isaacs et al. (2018) demonstrated in their Comprehensibility Scale, the position of hesitation markers, which includes pauses, significantly influences listener perception, suggesting that a more fine-grained analysis of pause location is essential for measuring L2 fluency.

Pause ratios across different syntactic contexts have begun to be addressed in recent studies. Metrics such as between-clause (BC) and within-clause (WC) pause ratios have been proposed (Kahng, 2018; Suzuki & Kormos, 2020; Hsieh et al., 2019), as well as more fine-grained phrase-level ratios (Kallio et al., 2022; Coulange et al., 2024b).

Instead of a categorical approach based on constituent type, we propose to consider word-to-word syntactic distance as a continuous value. This allows word-level granularity to better characterize pause syntactic position. We operationalize this word-to-word syntactic distance as the number of closing and opening constituents between each word, regardless of constituent type.

In our preliminary study (Coulange & de Jong, 2025), we developed two categorical scores based on categorical constituent types (clause and phrase) and introduced a syntactic pause ratio (SPR) based on word-to-word syntactic distance. All three measures showed strong positive correlations with proficiency, with correlation coefficients following the pattern: clause-only < clause+phrase < SPR. Building on these findings, the present study investigates how effectively syntactic distance predicts pause position and duration compared to traditional clause and phrase information, and examines its contribution to explaining variance in L2 overall proficiency when combined with articulation rate and pause frequency.

2. Literature Review

2.1 Syntactic Structure Predicting L1 and L2 Pauses

The relationship between pauses and syntactic structure for L1 speech has been a subject of investigation for over five decades. Early work established distinctions

between "grammatical pauses" and "hesitation pauses" (Boomer & Dittman, 1962; Boomer, 1965; Goldman-Eisler, 1968; Ruder & Jensen, 1972), also referred to as "structuring pauses" (Candea, 2000) and "disfluent pauses" (Fauth & Trouvain, 2018). One of the foundational studies in L1 speech was by Grosjean and Deschamps (1975), who used a corpus of BBC radio and analyzed 30 live interviews of minimum three minutes each. They found that 55% of all clauses were followed by a silent pause, with more pauses for higher-level clauses: 82% of final clauses, 71% of clauses followed by a coordinate, 60% of clauses followed by a non-relative clause, and 24% of clauses followed by a relative clause. Pauses occurring within clauses represented 45% of all pauses, and those occurring within phrases 26%.

More recently, Cao and Chen (2019) analyzed pause placement in prepared speech from what they called "successful speakers" (both L1 and L2) using TED talks (public presentations on various topics), presidential speeches, and videos from famous youtubers. They mainly observed pauses between subordinate and main clauses, and between clauses generally, with no significant differences between these successful or high proficiency L1 and L2 speakers regarding pause placement and duration. The study also identified emphatic pauses at lower syntactic levels, such as "fought for // the nations."

From these studies, we can conclude that proficient speakers tend to avoid pausing within syntactic units. Turning to L2 speech, it has been hypothesized that as proficiency develops, L2 learners will use fewer and shorter pauses overall, but particularly within syntactic units. This hypothesis is based on the premise that the linguistic encoding processes of formulating and articulating in an L2 are less automatic (Kormos, 2006; Skehan et al., 2016). In L2 speech production, these processes require more conscious effort and attentional control, which leads them to run more slowly compared to L1 speech. In L1 speech production, conscious attention and control are usually required only for conceptualizing messages. Since conceptualization mainly happens between major syntactic units, pauses within units tend to diminish as learners become more proficient.

Empirical studies support this hypothesis, showing that compared to L1 speech, L2 speech contains more pauses, especially within syntactic units. Furthermore, as proficiency increases, learners' pause patterns become more similar to those of L1 speakers. In an early study on L2 fluency, Riggenbach (1991), adopting a micro-analytic approach, observed that pauses at junctures could be considered "fluent", whereas those within syntactic units would be "disfluent". Syntactic units in this type of research have been defined either as AS-Units (e.g., De Jong, 2016), clauses (e.g., Kahng, 2014; Riggenbach, 1991; Tavakoli et al., 2020), or as

phrases[1] (Kallio et al., 2022; Coulange et al., 2024b; Riazantseva, 2001). A recent meta-analysis found that as L2 proficiency develops, pauses "migrate" from within syntactic units to the boundaries (Yan et al., 2025). Finally, neural data corroborates these findings showing increased brain activity in conceptualization-related areas at end-clause pauses but not mid-clause pauses (Révész et al., 2024).

In her corpus study comparing native speakers of English to L1 German learners of English, Götz (2013) investigated differences between the two groups of speakers regarding multiple aspects of fluency, including silent pauses within clauses and silent pauses within phrases. The difference between native speakers and learners was most pronounced for pauses within phrases (Götz, 2013, p. 99 – 100). Another study combining the different types of syntactic boundaries is by Coulange et al. (2024b), who examined 176 French learners of L2 English at B1 and B2 levels, analyzing 11 hours of spontaneous argumentative speech containing 22,000 pauses (180 ms — 2 s threshold). The study computed between-clause, between-phrase, and within-phrase pause ratios, finding that B1 speakers produced significantly more within-phrase pauses compared to B2 speakers, with no significant difference for between-clause pauses. To mitigate overall syntactic complexity differences between proficiency groups, Coulange (2025) refined this analysis by normalizing pause ratios with the number of constituent boundaries of each type. The results showed that only within-phrase pause ratios remained significantly higher for lower proficiency speakers.

Regarding pause duration, research consistently demonstrates that higher syntactic boundaries correlate with longer pauses. In L1 French, Candea (2000) and Grosman et al. (2018) found this pattern across read, prepared, and spontaneous speech. Tauberer (2008) reported similar findings for L1 English, and de Jong (2016) for L2 Dutch, though detailed pause duration analyses remain limited in the literature.

2.2 Pause Position and L2 Fluency Perception

The relationship between syntactic pause placement and fluency perception has been investigated across multiple L2 contexts. Mirroring the studies showing that as L2 learners progress, pauses within syntactic units become scarcer, studies on fluency perception show that pauses within units tend to be more harshly penal-

1. Some authors, such as Riazantseva (2001) or Götz (2013) use the word "constituent" to refer to phrases. Here, we will follow the classical syntactic theory and define "constituent" as any linguistic unit that functions as a single coherent entity within a hierarchical syntactic structure. Therefore, both clauses and phrases are syntactic constituents, with phrases being lower-level constituents than clauses. See Bhatt (2008) for further definitions.

ized than those at syntactic boundaries. For instance, Kahng (2018) examined L1 English raters' evaluation of spontaneous speech recordings from 37 Korean learners of English and 3 native speakers. Recordings were about 20 seconds each in duration, and rating was done globally using a 9-point Likert scale. Multiple linear regression analysis revealed that within-clause pause ratio explained 54% of variance in fluency perception, while between-clause pause ratio contributed only an additional 6%. Neither overall pause frequency nor mean pause duration improved the model. A follow-up experiment demonstrated that artificially inserted pauses significantly decreased fluency ratings, with stronger effects for within-clause than between-clause pauses.

These findings were replicated by Suzuki and Kormos (2020), who asked 10 L1 English raters to evaluate 40 Japanese learners of English (A2-C1) producing argumentative speech. Ratings were given on a 9-point scale for both fluency and comprehensibility. Pauses were categorized as end-. The authors found that fluency judgment was most strongly associated with mid-clause pause frequency, and that fluency and comprehensibility judgments were highly correlated.

Kallio et al. (2022) extended this research to L2 Finnish. Sixteen expert L1 Finnish raters evaluated 200 L2 learners on 4-point fluency and 7-point proficiency scales. Multiple regression analysis examined between- and within-clause, between- and within-phrase, as well as within-word pauses (non-terminated words). Within-phrase and between-phrase pause ratios showed strong correlations with fluency perception and proficiency judgment, whereas between-clause ratios showed only a weak correlation for fluency and no significant correlation for proficiency.

Finally, Coulange et al. (2024c) employed a dynamic rating methodology adapted from Nagle et al. (2019) with 60 L1 English raters evaluating 16 recordings (26–66 seconds each) from French learners of English (B1 and B2 levels). The study demonstrated a significant increase in perceived listener effort following within-phrase pauses, while effort tended to decrease after between-clause pauses, providing real-time evidence for the differential impact of pause placement on comprehensibility.

2.3 Related Work

The studies described above have shown that syntactic information about phrase and clause boundaries is related to both the occurrence and duration of pauses in speech, for both L1 and L2 speakers. Moreover, research on the perception of fluency in L2 speech has demonstrated that pauses occurring at locations other than syntactic boundaries tend to result in lower perceived fluency. These studies have in common that the syntactic information is operationalized as categorical:

pauses occur either at phrase/clause boundaries or within phrases/clauses. However, syntactic structure at the level of word-to-word transitions is inherently gradient. As Révész et al. (2024, p. 1200) mention in their limitations: *"it would be worthwhile to distinguish pause locations in terms of more specific syntactic constituents (e.g., different types of phrases)."* Below, we describe two earlier studies that have incorporated this gradient perspective to investigate the relationship between pauses and syntactic information.

In a study on read speech, Schweitzer and Haase (2000) compared part-of-speech tagging and syntactic trees to predict prosodic boundaries in a corpus of 67 minutes of radio news. Both methods proved to be promising. In the syntactic-tree analysis, they considered the relative count of closing brackets of the current word to the number of opening brackets of the previous word. This relative count proved to predict prosodic phrase boundaries in read speech, especially when the relative count turned out to be positive (i.e., when the count of closing brackets was higher than the number of opening brackets of the previous word). Tauberer (2008) investigated spontaneous speech and used decision trees to predict pause position and duration based on syntactic information in telephone conversations from the Switchboard corpus. He compared 12 parameters including timing, part-of-speech tags, and constituent information. The study introduced a "depth of boundary" measure, defined as the larger of the number of close-brackets or open-brackets at word boundaries, using bracket notation from constituency analysis (see Section 3.2 for further explanation). Notably, Tauberer found that the information about preceding constituent duration and following constituent word count predicted pause position and duration almost as accurately as the full 12-parameter model (F-score 78% vs. 79%).

To investigate the relationship between syntactic parsing and prosodic phrasing cross-linguistically, Kuang et al. (2022) also used closing and opening brackets as a proxy for the relative strength of the syntactic boundaries between adjacent words. In corpora of English and Mandarin L1 read speech, they compared the relation between closing and opening brackets and a conglomerate of acoustic features capturing pauses, duration cues, F0, energy, and voice quality. They found that the number of closing brackets had a stronger relation with the features capturing prosodic phrasing compared to the number of opening brackets. They also showed that of all acoustic features, pausing seemed to show the strongest relation to the strength of the syntactic boundaries.

2.4 Research Questions

Building on this foundation, the present study addresses two key research questions:

- **RQ1:** Is word-to-word syntactic distance a better predictor of pause position and duration than constituent types (clause and phrase)?
- **RQ2:** How much variance in L2 proficiency can a syntactic pause ratio based on continuous syntactic distance explain when combined with traditional fluency measures?

These questions aim to advance our understanding of pause placement in L2 speech and contribute to more nuanced fluency assessment methodologies. Our operationalization of syntactic distance is similar to the "depth of boundary measure" as evaluated by Tauberer (2008), the count of closing and opening brackets by Kuang et al. (2022), and the relative count of closing brackets of the current word to the number of opening brackets of the previous word in Schweitzer and Haase (2000). But rather than choosing the largest of the two options of closing and opening brackets (as in Tauberer, 2008) or specifically comparing closing and opening brackets (Kuang et al., 2022), we combine the two numbers of closing brackets and opening brackets between the current word to the next. We hypothesise that in spontaneous speech, the occurrence and duration of a pause can be both predicted by the speaker finishing a large constituent (to structure speech and aid the listener) as well as by the speaker embarking on a new large constituent (as larger and deeper embedded constituents call for more speech planning).

We therefore hypothesize that the continuous measure of syntactic distance better predicts pause position and duration compared to the categorical phrase and clause constituent boundaries (RQ1) and that information about pause placement using syntactic distances significantly adds explained variance when combined with traditional fluency measures (RQ2).

3. Method

3.1 Data

This study utilizes data from two complementary corpora available through ORTOLANG: the CLES-JP corpus (Japanese learners of English) and the CLES-EN corpus (native English speakers) (Coulange et al., 2024a). Speech recordings have been transcribed and annotated using the Pause and Lexical Stress Processing Pipeline (PLSPP, Coulange et al., 2024b).

CLES-JP Corpus[2]

The CLES-JP corpus comprises 4 hours of spontaneous L2 English speech from 30 Japanese university students (15 recordings) recorded at Waseda University, Tokyo (December 2023) and Doshisha University, Kyoto (February 2024). The corpus features 10 to 15-minute role-play discussions where pairs of students engage in argumentative discourse on contentious topics. Each participant was assigned a specific role (advocating for or against the topic) and allowed up to 10 minutes of preparation, with note-taking permitted but reading during conversation prohibited. The task required participants to negotiate, exchange viewpoints, and work toward compromise, mirroring the format of the CLES English certification exam[3] (CLES B2). Participants ranged from CEFR levels B1 to C1, based on equivalence with official certification tests such as TOEFL or IELTS. One recording featuring a Japanese-L1 English teacher was excluded from our analyses.

CLES-EN Corpus[4]

The CLES-EN corpus contains 2 hours of spontaneous L1 English speech from 14 US university students (7 recordings) recorded at Doshisha University, Kyoto (February 2024). The corpus follows the same task format as CLES-JP, with pairs of native English speakers engaging in the same argumentative role-play discussions. This corpus serves as the native speaker baseline for comparison with L2 productions.

All speech data was processed using the PLSPP, a comprehensive framework specifically designed for annotating pause and lexical stress patterns in spontaneous non-native English speech. The pipeline is optimized for batch processing of corpus files containing multiple speakers and produces detailed linguistic annotations for each audio file. No manual intervention was performed on the annotated data.

The PLSPP processing sequence relevant for our study includes the following steps:

1. **Neural speaker diarization**: Implemented using Pyannote.audio (Bredin, 2023) to detect voice activity and attribute segments to individual speakers. Consecutive segments from the same speaker were merged when separated by less than 1 second of silence, resulting in monospeaker continuous speech

2. Corpus CLES-JP: https://hdl.handle.net/11403/cles-jp_corpus/v1

3. CLES Certification: https://www.certification-cles.fr/english/

4. Corpus CLES-EN: https://hdl.handle.net/11403/cles-spontaneous-english/v1

segments. Only resulting speech utterances longer than 8 seconds were analyzed, including those containing overlapping speech, laughter and noise.
2. **Automatic speech recognition and word-level alignment**: Transcription was performed using WhisperX (Bain et al., 2023) for high-precision transcription and temporal alignment (Whisper model: base.en, forced alignment done using Wav2Vec2.0, Baevski et al., 2020).
3. **Syntactic analyses**: A dual approach combining SpaCy (Montani et al., 2023, model en_core_web_md) for morphological tagging and Berkeley Neural Parser (Kitaev, 2019, model benepar_en3) for constituency parsing.

Note that speech utterances are segmented based on pauses and speaker turns rather than grammatical sentence boundaries, and syntactic parsing is applied to these segments regardless of their sentential completeness. The pipeline generates a csv table listing all inter-word intervals with syntactic position and duration information. These intervals may be silent or filled with lengthening, filler particles such as "uh"/"uhm" or truncated words, which are typically not transcribed by the speech recognition model. For this study, pauses were extracted from this table using the duration threshold criteria described below.

The final dataset consisted of 39,957 words and 9,241 pauses from 43 speakers distributed across proficiency levels as follows: 5 B1, 15 B2, 9 C1 and 14 native English speakers.

3.2 Pause Prediction

The minimum pause duration threshold was set following Heldner and Edlund's (2010) recommendations: a minimum of 180 ms to exclude coarticulation (preburst) silences. Moreover, a maximum duration threshold of 2 seconds was also set to avoid excessively long pauses that might result from alignment errors. To approximate normal distribution for statistical analysis, all pause durations were log-transformed.

Syntactic constituents were classified into three categories "clause", "phrase" and "word," based on the larger ending or starting constituent at each word boundary. This classification follows Penn Treebank II constituent tags (Bies et al., 1995):

- **Clause level**: Boundaries where any ending or starting constituent is a clause-level tag (e.g., declarative clauses, clauses introduced by a subordinating conjunction or a *wh-word*, etc.)
- **Phrase level**: Boundaries where any ending or starting constituent is a phrase-level tag (e.g., noun phrases, verb phrases, prepositional phrases, etc.)
- **Word level**: All remaining constituent boundaries

Word-to-word syntactic distance was computed as the sum of closing and opening constituents at each word boundary (i.e., the number of closing and opening nodes in the constituency tree). This yields integer values ranging from 0 to infinity. Figure 1 shows an example of constituency analysis and syntactic distance computation, with details of the computation given in Table 1. Figure 2 shows the same example within its full utterance context.

Table 1. Details of word-to-word syntactic distance computation for the example in Figure 1

Word-to-word boundaries	Ending constituents	Starting constituents	Syntactic Distance (sum of ending and starting constituents)
you can	NP	VP	2
can get	–	VP	1
get some	–	NP	1
some information	–	–	0
information and	NP, VP, VP, S	–	4
and then	–	–	0
then you	–	S, NP	2
you can	NP	VP	2
can expand	–	VP	1
expand from	–	PP	1
from that	–	NP	1
that [end]	NP, PP, VP, VP, S, S	–	6

To normalize the distribution, a log-transformation was applied after adding 1 to all values, resulting in final values ranging from 0 to 4.64.

The following speaker-level measures were computed:

- **Articulation Rate (AR)**: Number of words divided by speech duration without pauses
- **Pause Frequency (PF)**: Number of pauses divided by number of words
- **Mean Pause Duration (PD)**
  S_clause: Weighted sum of between-clause pauses (+1) and within-clause pauses (−1), divided by total number of pauses
- **S_phrase**: Weighted sum of between-clause (+1), between-phrase (+0.5), and within-phrase (−1) pauses, divided by total number of pauses
- **Syntactic Pause Ratio (SPR)**: Mean syntactic distance of word boundaries with a pause divided by mean syntactic distance of all word boundaries

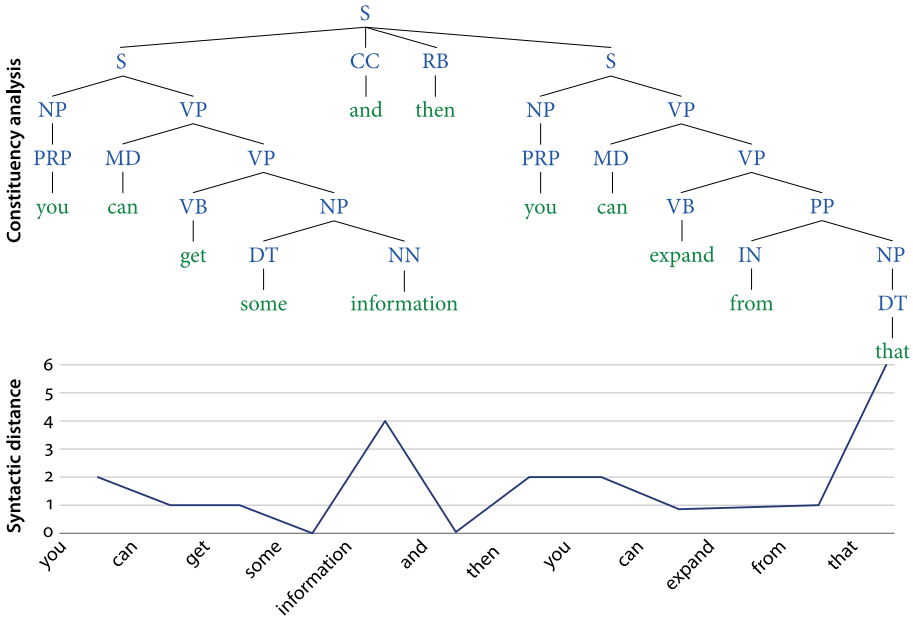

Figure 1. Constituency analysis and word-to-word syntactic distance computation of the utterance "you can get some information and then you can expand from that", extracted from the file doshisha2024_001_JNS_03A-03B_A_12 from the CLES-JP corpus; Bracket notation: [S [S [NP [PRP you]] [VP [MD can] [VP [VB get] [NP [DT some] [NN information]]]]] [CC and] [S [ADVP [RB then]] [NP [PRP you]] [VP [MD can] [VP [VB expand] [PP [IN from] [NP [DT that]]]]]]], annotated with Berkeley Neural Parser (Kitaev et al., 2019))

3.3 Statistical Analyses

Pause Occurrence Prediction

Generalized linear mixed models (glmer function from the lme4 R package) were employed to predict pause occurrence as a binary outcome after each word (pause vs. no pause; $N = 39{,}957$). Fixed effects included proficiency level and either constituent level or log syntactic distance, with speaker as a random effect to account for individual baseline pausing tendencies. Native speakers served as the reference level for proficiency comparisons.

Pause Duration Prediction

Linear mixed models (lmer function from the lme4 R package) were used to predict log-transformed pause duration ($N = 9{,}241$) using the same fixed effects structure and random effect specification as the generalized linear mixed models for pause occurrence.

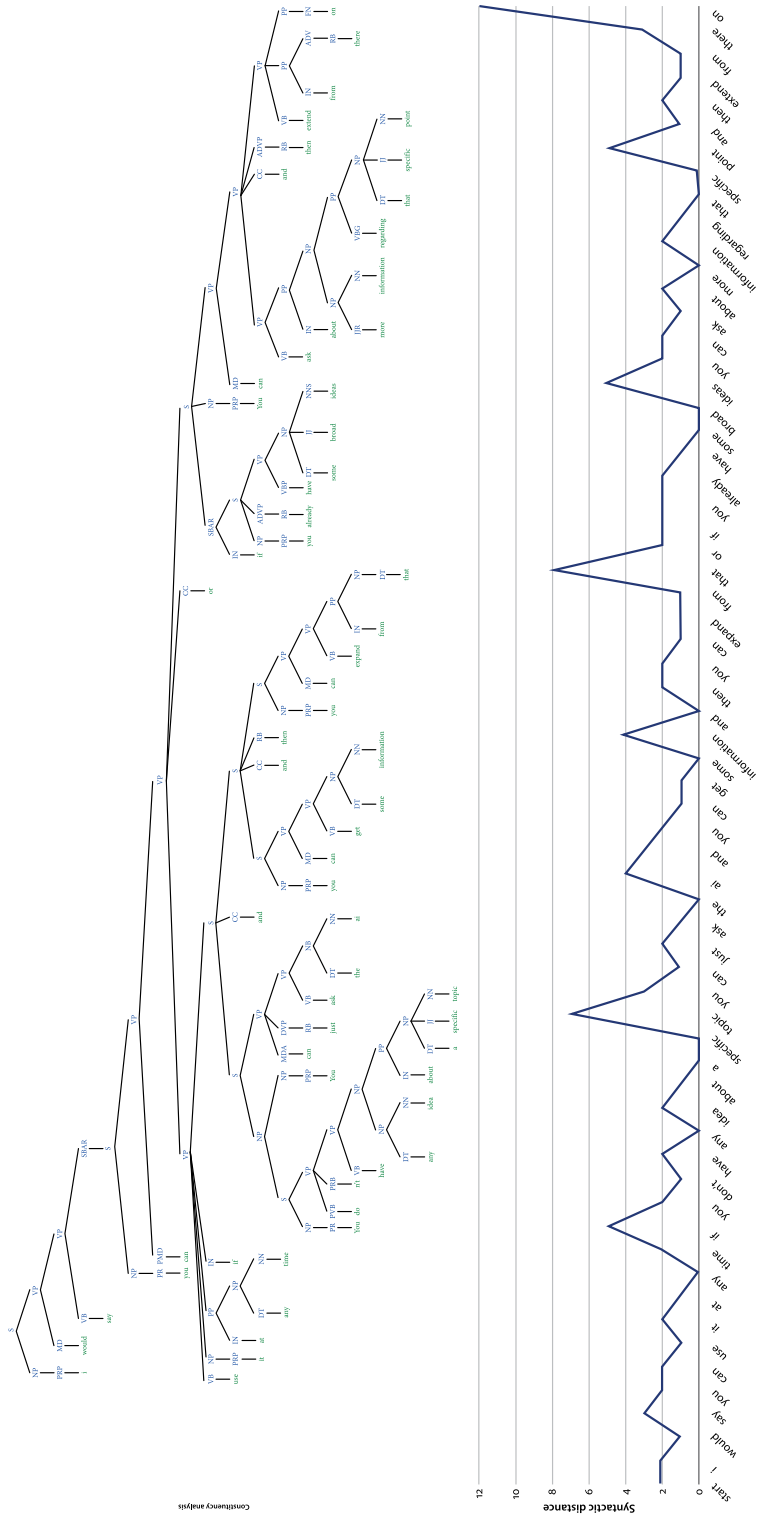

Figure 2. Constituency analysis and word-to-word syntactic distance computation on a full example (same file as Figure 1)

Proficiency Classification

Multinomial logistic regression (multinom from nnet R package) was employed to classify speakers ($N = 43$) into CEFR proficiency levels (categorical dependent variable) predicted by fluency measures.

Two baseline models were established:

1. Articulation Rate + Pause Frequency
2. Articulation Rate + Pause Frequency + Pause Duration

These baselines were then compared against augmented models incorporating either:

- Clause boundary information (S_clause)
- Clause and phrase information (S_phrase)
- Continuous syntactic distance (SPR)

Model performance was assessed using three complementary metrics computed with the performance R package. The Akaike Information Criterion (AIC) evaluates model quality by balancing goodness of fit against model complexity, with lower values indicating better models. McFadden's R^2 measures the proportion of variance in the dependent variable explained by the model, with values approaching 1 indicating better explanatory power. Root Mean Square Error (RMSE) quantifies the average magnitude of prediction errors, with lower values representing more accurate predictions. Together, these metrics help determine which syntactic measures best predict pause occurrence and duration, and which combination of measures best explains variance in proficiency classification.

4. Results

4.1 Descriptive Statistics

To characterize the speech fluency patterns in our dataset, we examined several key measures across proficiency levels: articulation rate, pause frequency, pause mean duration and pause locations based on constituent boundary types (clauses and phrases) and syntactic pause ratio. Descriptive statistics for fluency measures across proficiency levels are presented in Table 2. Articulation rate showed a clear progression from lower to higher proficiency levels, with B1 learners producing speech at 1.31 words per second, increasing through B2 (1.94) and C1 (3.25) to native speakers (3.79). Conversely, pause frequency and mean duration demonstrated the opposite pattern, decreasing from B1 (0.447 pauses per word and 0.784

seconds on average) to native speakers (0.176 and 0.572), indicating faster speech with fewer and shorter pauses at higher proficiency levels (see Figure 3).

The syntactic pause measures (S_clause and S_phrase) showed gradual increases toward native-like values, with S_clause progressing from −0.373 (B1) to −0.034 (native) and S_phrase from −0.063 (B1) to 0.429 (native). Similarly, the SPR measure increased across proficiency levels, moving from negative values in learner groups (B1: −0.565, B2: −0.454, C1: −0.244) to positive values for native speakers (0.159). These latter three measures all indicate that pauses tend to be placed more often at higher syntactic boundaries by speakers with higher proficiency (see Figure 4).

Table 2. Mean and standard deviation (SD) for each measure across each proficiency group

Measure Mean (SD)	B1	B2	C1	Native
Articulation Rate	1.31 (0.263)	1.94 (0.478)	3.25 (0.505)	3.79 (0.382)
Pause Frequency	0.447 (0.062)	0.345 (0.066)	0.239 (0.046)	0.176 (0.046)
Mean Pause Duration	0.784 (0.045)	0.662 (0.062)	0.589 (0.070)	0.572 (0.065)
S_clause	−0.373 (0.451)	−0.286 (0.443)	−0.133 (0.373)	−0.034 (0.316)
S_phrase	−0.063 (0.605)	−0.019 (0.501)	0.084 (0.457)	0.429 (0.225)
SPR	−0.565 (0.141)	−0.454 (0.272)	−0.244 (0.192)	0.159 (0.300)

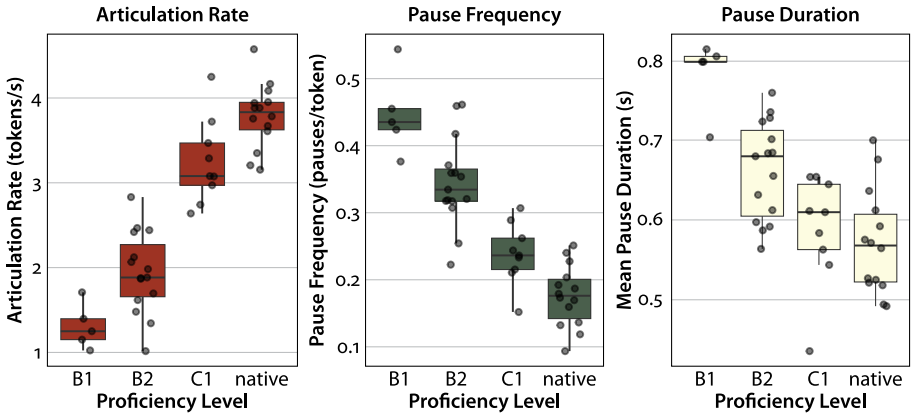

Figure 3. Articulation rate (left, in word tokens per second), overall pause frequency (middle, in pauses per word token) and mean pause duration (right, in seconds) per speaker in the dataset

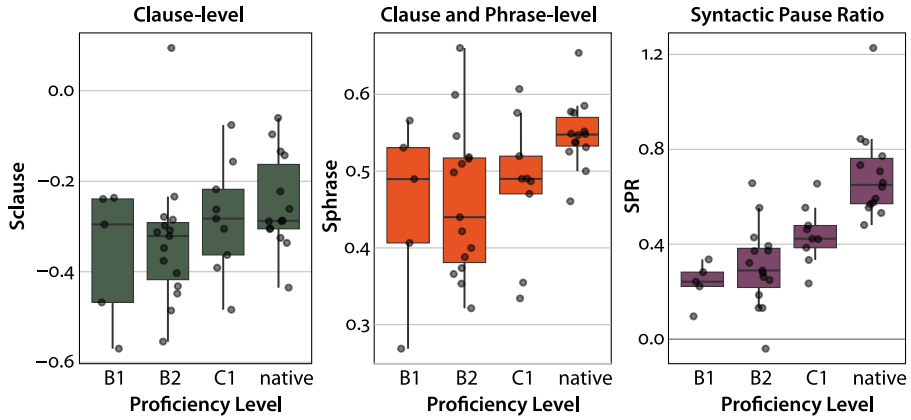

Figure 4. Three different scores (pause-level, pause- and phrase-level, and syntactic pause ratio) of pause placement per speaker

4.2 Pause Occurrence Prediction

Pause Occurrence and Clause/Phrase Boundaries

The first model examined pause occurrence using categorical syntactic boundaries (within-phrase, between-phrase, and between-clause) across different proficiency levels. The generalized linear mixed-effects model with binomial distribution revealed significant main effects for both boundary type and proficiency level.

Compared to within-phrase positions (reference level), pauses were significantly more likely to occur at phrase boundaries ($\beta = 0.62$, $z = 9.38$, $p < .001$) and even more so at clause boundaries ($\beta = 1.34$, $z = 19.42$, $p < .001$). This confirms the expected hierarchical pattern where syntactic boundaries of higher order are associated with increased pause likelihood.

Proficiency level also significantly predicted pause occurrence, with lower proficiency speakers showing higher overall pause probabilities. Relative to native speakers, all non-native groups demonstrated significantly higher pause probabilities: C1 speakers ($\beta = 0.66$, $z = 4.52$, $p < .001$), B2 speakers ($\beta = 1.17$, $z = 8.70$, $p < .001$), and B1 speakers ($\beta = 1.56$, $z = 7.79$, $p < .001$).

Importantly, the model revealed a significant interaction between boundary type and proficiency for higher proficiency levels. Compared to native speakers, C1 and B2 speakers showed reduced sensitivity to phrase boundaries (C1: $\beta = -0.31$, $p < .001$; B2: $\beta = -0.20$, $p < .05$) and clause boundaries (C1: $\beta = -0.25$, $p < .05$; B2: $\beta = -0.34$, $p < .01$), while no significant interaction was observed with B1 speakers. This indicates that as proficiency increases, pause placement tends to concentrate on higher-level syntactic boundaries.

Pause Occurrence and Syntactic Distance

The second model examined pause prediction using syntactic distance as a continuous predictor across proficiency levels.

The model revealed a strong main effect of syntactic distance on pause likelihood ($\beta = 0.96$, $z = 29.20$, $p < .001$). The positive coefficient indicates that as syntactic distance between adjacent words increases, the probability of a pause occurring also increases significantly. Consistent with the previous model, proficiency level showed significant main effects, with lower proficiency speakers exhibiting higher overall pause rates compared to native speakers: C1 speakers ($\beta = 0.65$, $z = 4.82$, $p < .001$), B2 speakers ($\beta = 1.30$, $z = 10.60$, $p < .001$), and B1 speakers ($\beta = 1.66$, $z = 9.04$, $p < .001$).

The model also revealed a significant interaction between syntactic distance and proficiency level. All learner groups showed reduced sensitivity to syntactic distance compared to native speakers: C1 speakers ($\beta = -0.21$, $z = -4.09$, $p < .001$), B2 speakers ($\beta = -0.30$, $z = -5.62$, $p < .001$), and B1 speakers ($\beta = -0.23$, $z = -2.31$, $p < .05$).

This model demonstrated improved fit compared to the categorical boundary model, with a lower AIC (see Table 3), indicating better predictive performance. Figure 5 shows the probability of observing a pause at different syntactic levels for each proficiency group.

Table 3. Comparison between the two pause prediction models. The syntactic distance model shows superior performance across all metrics, with a substantially lower AIC (ΔAIC = 556.1), higher explanatory power (marginal $R^2 = 0.116$ vs 0.092), and lower prediction error (RMSE = 0.402 vs 0.406)

Model	AIC	Marginal R^2	RMSE
Clause/Phrase boundaries	40522.0	0.092	0.406
Syntactic distance	39965.9	0.116	0.402

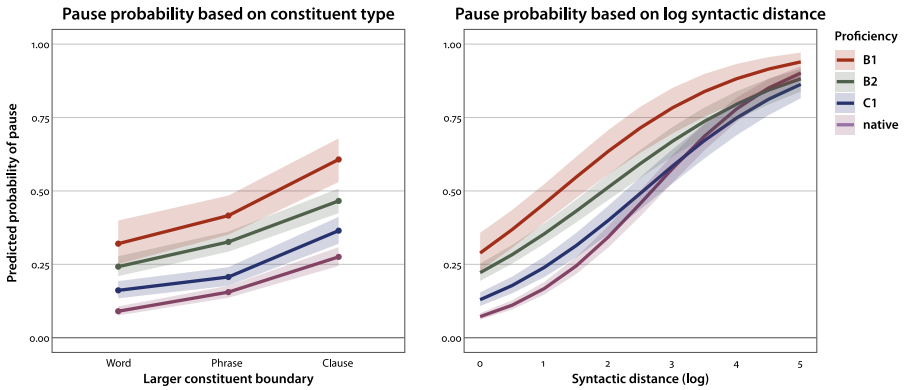

Figure 5. Predicted probabilities of pause for each model with confidence intervals

4.3 Pause Duration Prediction

Pause Duration and Clause/Phrase Boundaries

The linear mixed-effects model examining log-transformed pause duration revealed that syntactic boundary type significantly influenced pause length. Relative to within-phrase pauses (reference level), pauses at phrase boundaries were significantly longer ($\beta = 0.11$, $t = 3.12$), and pauses at clause boundaries were even longer ($\beta = 0.19$, $t = 5.12$). This hierarchical pattern indicates that pauses become progressively longer at higher-level syntactic boundaries.

Proficiency level also affected pause duration, with lower proficiency speakers producing longer pauses overall. Compared to native speakers, all non-native groups showed longer pause durations: C1 speakers ($\beta = 0.10$, $t = 1.70$), B2 speakers ($\beta = 0.13$, $t = 2.37$), and B1 speakers ($\beta = 0.25$, $t = 3.10$). The effect was most pronounced for B1 speakers, suggesting that lower proficiency is associated with substantially longer pauses.

The interaction effect revealed some variation in how proficiency modulated the relationship between boundary type and pause duration. C1 speakers showed reduced sensitivity to phrase boundaries ($\beta = -0.11$, $t = -2.12$), producing relatively shorter pauses at these locations compared to native speakers. B2 speakers also showed a similar but non-significant trend for phrase boundaries ($\beta = -0.09$, $t = -1.78$).

Pause Duration and Syntactic Distance

The linear mixed-effects model examining log-transformed pause duration using syntactic distance as a continuous predictor showed that syntactic distance significantly predicted pause length ($\beta = 0.08$, $t = 5.80$). As shown with the model based on clause and phrase boundaries, this indicates that speakers tend to modulate pause duration based on the degree of syntactic complexity at pause locations.

The main effects of proficiency level were smaller in this model compared to the categorical boundary model. While B1 speakers still showed significantly longer pauses than native speakers ($\beta = 0.17$, $t = 2.62$), the effects for C1 ($\beta = 0.04$, $t = 0.87$) and B2 speakers ($\beta = 0.04$, $t = 0.88$) were minimal and non-significant.

The interaction effect revealed how different proficiency groups respond to syntactic distance. While C1 speakers showed no significant modulation of the syntactic distance effect ($\beta = -0.02$, $t = -1.00$), both B2 ($\beta = 0.07$, $t = 3.10$) and B1 speakers ($\beta = 0.11$, $t = 2.99$) showed significantly stronger sensitivity to syntactic distance compared to native speakers. This indicates that even though lower proficiency speakers tend to make longer pauses at lower syntactic level than more advanced learners and native speakers, they still tend to make their longest pauses at high syntactic boundaries.

The two models show identical performance for predicting pause duration, with negligible differences in AIC, marginal R^2, and RMSE (see Table 4). Unlike the pause prediction models, neither approach demonstrates a clear advantage for modeling pause duration. Figure 6 gives a visual representation of pause duration prediction for different syntactic boundaries with both models.

Table 4. Comparison between the two pause duration prediction models

Model	AIC	Marginal R^2	RMSE
Clause/Phrase boundaries	130,000	0.036	0.583
Syntactic distance	130,000	0.033	0.584

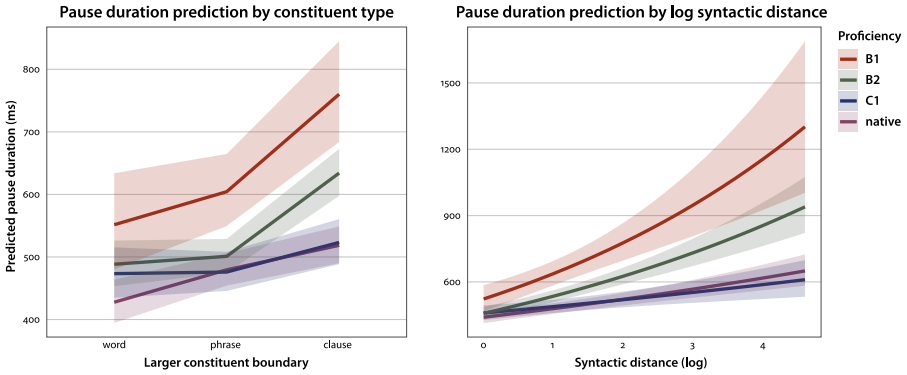

Figure 6. Predicted pause duration for each model with confidence intervals

4.4 Overall Proficiency Prediction

To assess the contribution of syntactic pause measures to proficiency classification, we compared multinomial logistic regression models with varying combinations of fluency predictors. Table 5 summarises each model performance.

Baseline Model Performance

The baseline model incorporating articulation rate and pause frequency achieved an R^2 of 0.654, explaining 65.4% of the variance in proficiency classification. Contrary to expectations, adding pause duration to this baseline model did not improve performance substantially ($R^2 = 0.659$, $\Delta AIC = +5.4$), and resulted in a higher AIC, indicating poorer model fit. Based on these results, pause duration was excluded from subsequent analyses, and the two-predictor baseline (articulation rate + pause frequency) was used for all further comparisons.

Categorical Syntactic Boundary Measures

Models incorporating categorical syntactic boundary information showed progressive improvements over the baseline. The model including clause boundary information (S_clause) increased explanatory power to $R^2 = 0.768$ ($\Delta AIC = -6.8$), while the model incorporating both clause and phrase boundary information (S_phrase) achieved $R^2 = 0.810$ ($\Delta AIC = -11.6$). Both models demonstrated substantially better fit than the baseline, with lower AIC values and reduced prediction error (RMSE = 0.221 and 0.187, respectively).

Continuous Syntactic Distance

The model incorporating the syntactic pause ratio (SPR) achieved the best performance across all metrics, with $R^2 = 0.845$, explaining 84.5% of the variance in proficiency classification. This model showed the lowest AIC (41.4) and prediction error (RMSE = 0.172).

Collinearity Assessment

Analysis of variance inflation factors confirmed our methodological decision to test syntactic measures separately. When S_clause, S_phrase, and SPR were included together, high collinearity was detected, with S_clause showing the highest VIF (11.18), followed by S_phrase (7.98) and SPR (4.43). This multicollinearity would compromise model interpretability and stability, justifying our approach of testing these measures individually.

The results demonstrate that continuous syntactic distance measures (SPR) provide the most robust contribution to proficiency classification, outperforming both traditional fluency metrics and categorical boundary-based approaches. In Table 5, model weights represent the relative probability that each model is the best approximating model among the candidate set, calculated using AIC differences (Burnham & Anderson, 2002). The SPR model obtains the higher weight (0.870), indicating better relative performance. Figure 7 gives a visual comparison of AIC and R^2 for each model.

5. Discussion

The present study investigated the role of syntactic complexity in L2 speech fluency assessment, proposing a continuous measure of pause placement based on syntactic distance between adjacent words. We conducted three main analyses using a corpus of spontaneous English speech produced by Japanese learners across three proficiency levels (B1, B2, C1) and native speakers.

Table 5. Proficiency classification model performance. AR = Articulation Rate, PF = Pause Frequency, PD = Pause Duration, SPR = Syntactic Pause Ratio. ΔAIC calculated relative to the baseline model

Model	AIC	ΔAIC	R^2	RMSE	Model Weight
AR + PF (Baseline)	57.0	0.0	0.654	0.272	< 0.001
AR + PF + PD	62.4	+5.4	0.659	0.271	< 0.001
AR + PF + S_clause	50.2	−6.8	0.768	0.221	0.011
AR + PF + S_phrase	45.4	−11.6	0.810	0.187	0.119
AR + PF + SPR	41.4	−15.6	0.845	0.172	0.870

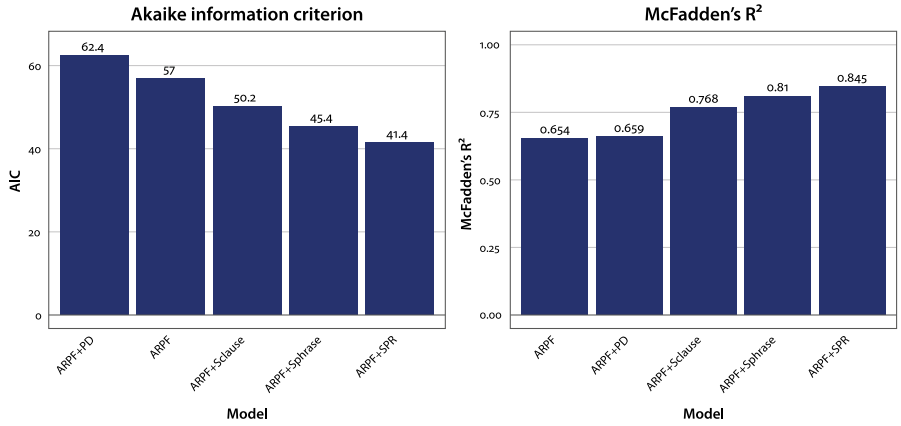

Figure 7. Akaike Information Criterion (left) and McFadden's R^2 (right) for each model

First, we examined pause occurrence using generalized linear mixed-effects models, comparing categorical syntactic boundaries (based either on clause, or clause and phrase boundaries) against continuous syntactic distance. The syntactic distance model substantially outperformed the categorical boundary model (AIC = 39965.9 vs. 40522.0, ΔAIC = 556.1), with higher explanatory power (marginal R^2 = 0.116 vs. 0.092) and lower prediction error (RMSE = 0.402 vs 0.406). Importantly, while all proficiency groups showed sensitivity to syntactic distance, native speakers demonstrated significantly greater sensitivity than learners, suggesting more sophisticated pause placement strategies.

Second, we investigated pause duration using linear mixed-effects models with the same predictor comparisons. Here, both categorical and continuous approaches showed similar performance, with neither model demonstrating a clear advantage. However, the syntactic distance model revealed that lower proficiency speakers showed stronger sensitivity to syntactic complexity in deter-

mining pause duration, indicating greater processing demands at syntactically complex locations.

Third, we evaluated proficiency classification using multinomial logistic regression. The model incorporating the syntactic pause ratio (SPR), our novel continuous measure, achieved the best performance, explaining 84.5% of the variance in proficiency classification compared to 65.4% for baseline measures of articulation rate and pause frequency (which is already quite high). This model outperformed all alternatives and exceeded categorical boundary-based approaches ($R^2 = 0.768$ for clause boundaries, $R^2 = 0.810$ for phrase boundaries).

These converging findings demonstrate that continuous syntactic distance measures provide finer-grained characterization of L2 speech fluency, particularly for pause prediction and proficiency classification, while revealing distinct patterns of syntactic sensitivity across proficiency levels.

5.1 Theoretical Implications

Our results challenge the prevailing approach in L2 fluency research that relies on discrete syntactic categories such as clause or phrase boundaries. The superior performance of continuous syntactic distance measures suggests that fluency is better characterized by gradations in syntactic complexity rather than binary (between or within clauses) or ternary categorical distinctions (between clauses, between phrases or within phrases). The continuous measure better captures the syntactic depth of constituent nesting, as well as gradations of syntactic breaks between different clause and phrase boundaries, which are not accounted for by constituent-type based categorical approaches, and yet clearly influences pause placement.

This gradual effect of syntactic distance would also be predicted by Kormos' (2006) theory on L2 speech production. As Révész et al. (2024) suggested, different types of phrase and clause boundaries may have distinct effects on speech production processes. In our operationalization, syntactic break size is treated as a continuous variable. Larger breaks increase the likelihood that speakers must engage in conceptual planning in addition to linguistic planning (formulating and articulating), which raise the chances of pausing. For L2 speakers, this effect is amplified, depending on proficiency: as breaks become larger, more conscious linguistic planning is required, leading to more frequent and potentially longer pauses.

Kuang et al. (2022) also reported that the prosodic realization of syntactic boundaries is gradient. Their research found that especially the depth of the syntactic boundary in terms of closing constituents was related to prosodic features, including pauses. Their finding is in line with Schweitzer and Haase (2000)' study

on German read speech, who also report a correlation between number of closing brackets (relative to the number of opening brackets of the previous word) and prosodic phrase boundaries. For spontaneous speech, however, we hypothesized that, in addition to the number of closing brackets, the depth of opening constituents to the next word should be considered, as the processing demands of upcoming large constituents together with constituents with deep embedding are likely to result in higher probabilities and longer durations of pauses. Tauberer (2008) indeed incorporated information about both preceding and upcoming constituents at any word boundary. His finding that the duration of the preceding constituent combined with the number of words of the following constituent at any word boundary had a similar predictive accuracy of pause occurrence compared to a feature set of 12 features (that included the biggest number of either closing or opening brackets in a syntactic parse) may not contradict our results. In fact, combining timing and word count of adjacent constituents is likely highly correlated with the syntactic distance measure used in the current study, which also integrates information about both closing and upcoming constituents. Our current operationalization of syntactic distance can be seen as a more parsimonious operationalization of *syntactic* structure, as it relies solely on syntactic information.

The differential sensitivity in our study to syntactic distance across proficiency levels offers insights into the cognitive mechanisms underlying L2 speech fluency. Native speakers demonstrated greater sensitivity to fine-grained syntactic complexity in pause placement, suggesting more efficient syntactic processing and integration. This pattern is consistent with theories of automaticity in L2 acquisition (Segalowitz, 2010), where advanced processing allows for more strategic use of pauses in relation to syntactic structure. Conversely, the stronger sensitivity to syntactic distance in pause duration among lower proficiency speakers likely reflects greater processing demands at syntactically complex locations, where speakers require additional time to plan and integrate upcoming linguistic material.

These findings contribute to a reconceptualization of L2 fluency that extends beyond traditional measures of speed and frequency. Rather than viewing pauses merely as disruptions to fluency, our results suggest that strategic pause placement in relation to syntactic structure is a hallmark of proficient speech. This perspective aligns with recent work emphasizing the importance of considering pause locations (de Jong, 2016; Kahng, 2018; Suzuki & Kormos, 2020; Kallio et al., 2022) and supports the notion that fluency encompasses not just temporal aspects but also the linguistic sophistication of pause use.

5.2 Methodological Contributions

The syntactic pause ratio (SPR) emerges as a robust, and intuitive measure that effectively combines pause behavior with syntactic structure. It is directly related to local predictions of pause probability and pause duration. Measures derived from Tauberer's (2008) results, which involve both the duration of the preceding constituent and the number of words of the upcoming constituent, may be less intuitive because they combine syntactic structure, number of words, and articulation rate. Unlike traditional approaches in L2 fluency research that treat syntactic structure in a categorical manner, SPR provides a continuous metric that better captures the relationship between pausing patterns and syntactic complexity. The measure's superior performance in proficiency classification, despite the additional challenges in automatic processing of L2 speech, demonstrates its potential as a reliable indicator of L2 fluency development.

Our comparison of continuous versus categorical measures reveals the advantages of gradient approaches for capturing linguistic phenomena. The substantial improvement in model fit when using syntactic distance underscores the limitations of categorical boundary-based methods. This finding has broader implications for computational linguistics and natural language processing, suggesting that continuous measures may be more effective for modeling human language behavior across various domains.

5.3 Practical Applications and Pedagogical Implications

These findings have direct implications for automated language assessment systems. The robust performance of syntactic distance measures suggests that incorporating syntactic complexity into fluency assessment algorithms could significantly improve their accuracy and validity. The continuous nature of these measures makes them particularly suitable for computational implementation, offering a more sophisticated alternative to current approaches that rely primarily on temporal measures.

From a pedagogical perspective, these results suggest that fluency instruction should address not only the temporal aspects of speech but also the strategic use of pauses in relation to syntactic structure. Teaching learners to recognize and utilize syntactic boundaries for pause placement may contribute to fluency patterns, making speech easier to process for the listener.

To facilitate the adoption of these methods by the broader research and pedagogical community, the complete speech processing and annotation code is made available through the open-source PLSPP pipeline,[5] which can be down-

5. PLSPP pipeline: https://gricad-gitlab.univ-grenoble-alpes.fr/lidilem/plspp

loaded and reused freely. The pipeline implements both categorical and continuous approaches to syntax-based pause annotation, enabling researchers to replicate and extend our analyses. For those without technical expertise, the same functionality is accessible through PLSPP Web Services,[6] a web interface that allows teachers and researchers to annotate their own speech data online with a GDPR-compliant service. These resources aim to bridge the gap between theoretical advances in L2 fluency research and practical applications in language assessment and instruction.

5.4 Limitations and Future Directions

Several limitations still need to be addressed in upcoming work. Our findings are based on Japanese learners of English, and cross-linguistic validation is necessary to establish the generalizability of the relationship between syntactic distance and pause patterns. Different L1-L2 combinations may yield different results due to varying degrees of syntactic transfer and processing strategies. Moreover, beyond structural factors, pausing strategies may also differ among cultures. For instance, while pauses are often avoided and interpreted as turn-change opportunities by English native speakers (Sacks, 1992; Fox et al., 1996), short pauses are quite frequent in Japanese and allow simultaneous reactions by the interlocutor without speech turn interruption (Maynard, 1989). Shigemitsu (2007) showed, for example, how using English pausing strategies in Japanese or Japanese pausing strategies in English can disturb conversations between English and Japanese native speakers.

While our sample size was sufficient for the current analyses, larger-scale studies would strengthen the robustness of these findings and allow for more fine-grained analyses of individual differences. The small number of B1 speakers constitutes a notable limitation that might explain the low significance of some results, such as S_clause and S_phrase scores. In previous work (Coulange, 2025), similar metrics were used to characterize the syntactic distribution of pauses among 170 French learners of English at CEFR B1 and B2 levels, showing a significant difference between both levels, with the larger difference occurring for the continuous syntactic distance-based score.

Another limitation is that our constituency parsing was based exclusively on the text transcript (with no prosody information) and that the model we used was primarily trained on written text. The precision of this syntactic analysis could be improved by using a speech-based syntactic parser. However, this would intro-

6. PLSPP Web Services: https://plspp.univ-grenoble-alpes.fr/

duce circularity because such parsers typically incorporate prosodic cues, including pauses, into their predictions.

Additionally, this study establishes associations rather than causal relationships between syntactic complexity and pause behavior. Future research should investigate whether training learners to attend to syntactic structure in pause placement leads to improvements in perceived fluency and overall proficiency ratings.

Several directions for future research emerge from these findings. Longitudinal studies tracking the same learners over time would provide insight into the developmental trajectory of syntactic pause placement strategies. Cross-linguistic studies examining learners from different L1 backgrounds would help establish the universality of these patterns versus language-specific effects. Finally, investigating these patterns in different speech contexts (e.g., spontaneous versus planned speech, different task types) would clarify the scope and stability of syntactic distance effects on pause behavior.

Conclusion

This study advances our understanding of L2 speech fluency by demonstrating the crucial role of syntactic complexity in pause placement. While classical approaches based on constituent boundary types (e.g., clauses and phrases) provide useful insights into syntax-pause relationships, this categorical approach captures syntactic complexity only to a limited extent. The continuous measure of syntactic distance adopted in this study, building on preliminary work, proved substantially more informative, robust, and effective for characterizing pause patterns in relation to syntactic structure.

By moving beyond simple temporal measures to incorporate the strategic use of pauses in relation to syntactic structure, we gain a more nuanced and theoretically grounded understanding of what constitutes fluent L2 speech. These findings have significant implications for both theoretical models of L2 fluency and practical applications in language assessment and instruction. They offer a more sophisticated framework for understanding and evaluating second language speech fluency, one that recognizes pause placement as a strategic linguistic behavior that supports discourse organization and listener comprehension, rather than merely disrupting temporal flow.

Funding

This work was made possible thanks to the FITInnovE 2024–2028 funding (Pôle Universitaire d'Innovation Grenoble Alpes).

Code

- Speech annotation was done with the Pause and Lexical Stress Processing Pipeline (Coulange et al., 2024b). Source code is available on GitLab at: https://gricad-gitlab.univ-grenoble-alpes.fr/lidilem/plspp
- Computation of fluency scores and models were done with R. Scripts are provided here: https://doi.org/10.17605/OSF.IO/KW49Q

References

Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. (2020). Wav2Vec 2.0: A framework for self-supervised learning of speech representations. *Advances in Neural Information Processing Systems*, *33*, 12449–12460.

Bain, M., Huh, J., Han, T., Zisserman, A. (2023). WhisperX: Time-Accurate Speech Transcription of Long-Form Audio. *Proc. Interspeech 2023*, 4489–4493.

Bies, A., Ferguson, M., Katz, K., & MacIntyre, R. (1995). Bracketing Guidelines For Treebank II Style Penn Treebank Project. University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-95-06-07. LINC LAB 281 https://repository.upenn.edu/handle/20.500.14332/6998

Bhatt, R. (2008). Pharse Structure rules, Tree rewriting and recursion. Amherst: UMASS.

Carnie, A. (2002). *Syntax: A Generative Introduction*. Wiley-Blackwell.

Boomer, D.S. (1965). Hesitation and grammatical encoding. *Language and Speech*, *8*(3), 148–158.

Boomer, D.S., & Dittmann, A.T. (1962). Hesitation pauses and juncture pauses in speech. *Language and Speech*, *5*(4), 215–220.

Bredin, H. (2023). pyannote.audio 2.1 speaker diarization pipeline: principle, benchmark, and recipe. *Interspeech 2023*, 1983–1987.

Burnham, K.P., & Anderson, D.R. (2002). Model selection and multimodel inference (2nd ed.; K.P. Burnham & D.R. Anderson, Eds.).

Candea, M. (2000). Contribution à l'étude des pauses silencieuses et des phénomènes dits '"d'hésitation"' en français oral spontané. *Etude sur un corpus de récits en classe de français* (Université de la Sorbonne nouvelle — Paris III). https://theses.hal.science/tel-00290143

Cao, Y., & Chen, H. (2019). World Englishes and Prosody: Evidence from the Successful Public Speakers. *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2048–2052.

Coulange, S. (2025). Évaluation automatique de la parole spontanée en anglais langue étrangère : le rôle des pauses et de l'accent lexical dans la compréhensibilité du locuteur. *Thèse de doctorat en Sciences du langage Spécialité Informatique*, dirigée par Monica Masperi, Solange Rossato, et Tsuneo Kato, Université Grenoble Alpes.

Coulange, S., de Jong, N.H. (2025). Measuring L2 Speech Fluency Based on Syntactic Distribution of Pauses. *Proc. 12th edition of the Disfluency in Spontaneous Speech Workshop (DiSS 2025)*, 37–41.

Coulange, S., Konishi, T., Sugahara, M., & Kato, T. (2024a). A corpus of spontaneous dialogues in L2 English by French and Japanese L1 speakers for automated assessment of fluency. *6th International Symposium on Learner Corpus Studies in Asia and the World (LCSAW6)*. https://hal.science/hal-04666131

Coulange, S., Kato, T., Rossato, S., & Masperi, M. (2024b). Enhancing language learners' comprehensibility through automated analysis of pause positions and syllable prominence. *Languages*, *9*(3), 78.

Coulange, S., Kato, T., Rossato, S., & Masperi, M. (2024c). Exploring impact of pausing and lexical stress patterns on L2 English comprehensibility in real time. *Interspeech 2024*, 1030–1034. Presented at the Interspeech 2024.

de Jong, N.H. (2016). 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.

Fauth, C., & Trouvain, J. (2018). Détails phonétiques dans la réalisation des pauses en Français : étude de parole lue en langue maternelle vs en langue étrangère. *Langages*, *211*(3), 81–95.

Fox, B.A., Hayashi, M., & Jasperson, R. (1996). Resources and repair : a cross-linguistic study of syntax and repair. In E. Ochs, E.A. Schegloff & S.A. Thompson (Éd.), *Interaction and Grammar (p. 185–237)*. Cambridge Univ. Press.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in Spontaneous Speech*. Academic Press Inc.

Götz, S. (2013). *Fluency in Native and Nonnative English Speech* (Vol. *53*). John Benjamins Publishing Company.

Grosjean, F., & Deschamps, A. (1975). Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica*, *31*(3–4), 144–184.

Grosjean, F., & Deschamps, A. (2009). Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica*, *31*(3–4), 144–184.

Grosman, I., Simon, A.C., & Degand, L. (2018). Variation de la durée des pauses silencieuses : impact de la syntaxe, du style de parole et des disfluences. *Langages*, *211*(3), 13–40.

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*(4), 555–568.

Hsieh, C.-N., Zechner, K., & Xi, X. (2019). Features measuring fluency and pronunciation. In *Automated Speaking Assessment* (pp. 101–122).

Isaacs, T., Trofimovich, P., & Foote, J.A. (2018). Developing a user-oriented second language comprehensibility scale for English-medium universities. *Language Testing*, *35*(2), 193–216.

Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, *39*(3), 569–591.

Kahng, J. (2014). Exploring Utterance and Cognitive Fluency of L1 and L2 English Speakers: Temporal Measures and Stimulated Recall: Utterance and Cognitive Fluency in L2. *Language Learning*, *64*(4), 809–854.

Kallio, H., Kuronen, M., & Koivusalo, L. (2022). The role of pause location in perceived fluency and proficiency in L2 Finnish. *Proc. ISAPh 2022, 4th International Symposium on Applied Phonetics*, 22–27.

Kitaev, N., Cao, S., & Klein, D. (2019). Multilingual Constituency Parsing with Self-Attention and Pre-Training. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3499–3505.

Kormos, J. (2006). *Speech Production and Second Language Acquisition*. Routledge.

Kuang, J., Chan, M. P. Y., Rhee, N., Liberman, M., Ding, H. (2022). The mapping between syntactic and prosodic phrasing in English and Mandarin. *Proc. Interspeech 2022*, 3443–3447.

Maynard, S. K. (1989). *Japanese conversation: Self-contextualization through structure and interactional management*. Praeger.

Montani, I., Honnibal, M., Honnibal, M., Boyd, A., Van Landeghem, S., & Peters, H. (2023). spaCy: Industrial-strength Natural Language Processing in Python.

Nagle, C., Trofimovich, P., & Bergeron, A. (2019). Toward a dynamic view of second language comprehensibility. *Studies in Second Language Acquisition*, *41*(04), 647–672.

Révész, A., Jeong, H., Suzuki, S., Cui, H., Matsuura, S., Saito, K., & Sugiura, M. (2024). Task-generated processes in second language speech production: Exploring the neural correlates of task complexity during silent pauses. *Studies in Second Language Acquisition*, *46*(4), 1179–1205.

Riazantseva, A. (2001). Second Language Proficiency and Pausing: A Study of Russian Speakers of English. *Studies in Second Language Acquisition*, *23*(4), 497–526.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, *14*(4), 423–441.

Ruder, K. F., & Jensen, P. J. (1972). Fluent and hesitation pauses as a function of syntactic complexity. *Journal of speech and hearing research*, *15*(1), 49–60.

Sacks, H. (1992). Lectures on Conversation (G. Jefferson, Ed.). Blackwell.

Segalowitz, N. (01 2010). *Cognitive bases of second language fluency*. New York and London: Routledge.

Schweitzer, A., & Haase, M. (2000). Zwei Ansätze zur syntaxgesteuerten Prosodiegenerierung. *KONVENS 2000 / Sprachkommunikation, Vorträge Der Gemeinsamen Veranstaltung 5. Konferenz Zur Verarbeitung Natürlicher Sprache (KONVENS), 6. ITG-Fachtagung "Sprachkommunikation,"* 197–202.

Shigemitsu, Y. (2007). A pause in conversation for Japanese native speakers : a case study of successful and unsuccessful conversation in terms of pause though intercultural communication. *Academic Report*, Tokyo Polytechnic University, *30*(2), 11–18. https://cir.nii.ac.jp/crid/1520290882531293440

Shea, C., & Leonard, K. (2019). Evaluating measures of pausing for second language fluency research. *Canadian Modern Language Review*, *75*(3), 216–235.

Skehan, P., Foster, P., & Shum, S. (2016). Ladders and Snakes in Second Language Fluency. *International Review of Applied Linguistics in Language Teaching*, *54*(2).

Suzuki, S., & Kormos, J. (2020). linguistic dimensions of comprehensibility and perceived fluency: an investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, *42*(1), 143–167.

Tauberer, J. (2008). Predicting intrasentential pauses: is syntactic structure useful? *Proc. Speech Prosody 2008*, 405–408.

Tavakoli, P., Nakatsuhara, F., & Hunter, A. (2020). Aspects of Fluency Across Assessed Levels of Speaking Proficiency. *The Modern Language Journal*, *104*(1), 169–191.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In *Planning and Task Performance in a Second Language* (pp. 239–273).

Yan, X., Lei, Y., & Pan, Y. (2025). Diving Deep Into the Relationship Between Speech Fluency and Second Language Proficiency: A Meta-Analysis. *Language Learning*, *75*(4), 1051–1090.

Model Outputs

Pause prediction based on boundary type (const_model)

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']

 Family: binomial ( logit )
Formula: is_pause ~ level * Proficiency + (1 | spk)
   Data: data
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e+05))

AIC     BIC     logLik   -2*log(L) df.resid
40522.0 40633.7 -20248.0 40496.0   39944

Scaled residuals:

Min     1Q      Median  3Q      Max
-1.4982 -0.5551 -0.4315 -0.2664 4.1896

Random effects:

Groups Name        Variance Std.Dev.
spk    (Intercept) 0.07714  0.2777

Number of obs: 39957, groups: spk, 43
Fixed effects:

                          Estimate Std. Error z value Pr(>|z|)
(Intercept)               -2.31046 0.09603    -24.060 < 2e-16 ***
levelphrase               0.61748  0.06586    9.376   < 2e-16 ***
levelclause               1.34301  0.06917    19.417  < 2e-16 ***
ProficiencyC1             0.66259  0.14645    4.524   6.05e-06 ***
ProficiencyB2             1.16926  0.13437    8.702   < 2e-16 ***
ProficiencyB1             1.55829  0.20008    7.788   6.79e-15 ***
levelphrase:ProficiencyC1 -0.31342 0.09415    -3.329  0.000871 ***
levelclause:ProficiencyC1 -0.25012 0.09993    -2.503  0.012320 *
```

```
levelphrase:ProficiencyB2 -0.20127 0.09477    -2.124  0.033679 *
levelclause:ProficiencyB2 -0.33834 0.10292    -3.288  0.001011 **
levelphrase:ProficiencyB1  0.20446 0.15479     1.321  0.186535
levelclause:ProficiencyB1 -0.15546 0.17373    -0.895  0.370896
```

Pause prediction based on syntactic distance (logsyntdist_model)

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']

 Family: binomial ( logit )
Formula: is_pause ~ logSyntDist * Proficiency + (1 | spk)
   Data: data
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e+05))

AIC     BIC     logLik   -2*log(L) df.resid
39965.9 40043.3 -19974.0 39947.9   39948

Scaled residuals:

Min     1Q      Median  3Q      Max
-2.3338 -0.5599 -0.4174 -0.2344 4.7542

Random effects:

Groups Name        Variance Std.Dev.
spk    (Intercept) 0.07923  0.2815

Number of obs: 39957, groups: spk, 43
Fixed effects:

                          Estimate Std. Error z value Pr(>|z|)
(Intercept)               -2.55855 0.08636    -29.628 < 2e-16 ***
logSyntDist               0.95520  0.03271    29.203  < 2e-16 ***
ProficiencyC1             0.65367  0.13567    4.818   1.45e-06 ***
ProficiencyB2             1.30252  0.12289    10.599  < 2e-16 ***
ProficiencyB1             1.65545  0.18305    9.044   < 2e-16 ***
logSyntDist:ProficiencyC1 -0.20597 0.05040     4.086  4.38e-05 ***
logSyntDist:ProficiencyB2  0.30156 0.05368     5.618  1.93e-08 ***
logSyntDist:ProficiencyB1  0.22620 0.09797    -2.309  0.0209 *
```

Pause duration prediction based on boundary types (constDUR_model)

```
Linear mixed model fit by REML ['lmerMod']
Formula: log(durationMS) ~ level * Proficiency + (1 | spk)
   Data: data[data$is_pause == "TRUE", ]

REML criterion at convergence: 16419.5
Scaled residuals:

Min      1Q       Median   3Q      Max
-2.34787 -0.78476 -0.03453 0.72667 2.63949

Random effects:

Groups   Name        Variance Std.Dev.
spk      (Intercept) 0.007508 0.08665
Residual             0.341625 0.58449

Number of obs: 9241, groups: spk, 43
Fixed effects:

                           Estimate Std. Error t value
(Intercept)                6.05835  0.04118    147.110
levelphrase                0.11430  0.03660    3.123
levelclause                0.19280  0.03766    5.120
ProficiencyC1              0.10145  0.05985    1.695
ProficiencyB2              0.13275  0.05606    2.368
ProficiencyB1              0.25435  0.08204    3.100
levelphrase:ProficiencyC1 -0.10882 0.05138    -2.118
levelclause:ProficiencyC1 -0.09248 0.05319    -1.739
levelphrase:ProficiencyB2 -0.08882 0.04999    -1.777
levelclause:ProficiencyB2 0.06833  0.05236    1.305
levelphrase:ProficiencyB1 -0.02285 0.07568    -0.302
levelclause:ProficiencyB1 0.12803  0.07987    1.603
```

Pause duration prediction based on syntactic distance (logsyntdistDUR_model)

```
Linear mixed model fit by REML ['lmerMod']
Formula: log(durationMS) ~ logSyntDist * Proficiency + (1 | spk)
   Data: data[data$is_pause == "TRUE", ]

REML criterion at convergence: 16426.6
```

```
Scaled residuals:

Min      1Q       Median   3Q      Max
-2.32508 -0.78899 -0.03045 0.71865 2.59155

Random effects:

Groups   Name        Variance Std.Dev.
spk      (Intercept) 0.007387 0.08595
Residual             0.342339 0.58510

Number of obs: 9241, groups: spk, 43
Fixed effects:

                           Estimate Std. Error t value
(Intercept)                6.08813  0.03117    195.299
logSyntDist                0.08447  0.01457    5.797
ProficiencyC1              0.04185  0.04832    0.866
ProficiencyB2              0.03894  0.04435    0.878
ProficiencyB1              0.17133  0.06534    2.622
logSyntDist:ProficiencyC1 -0.02279 0.02291    -0.995
logSyntDist:ProficiencyB2 0.07169  0.02309    3.104
logSyntDist:ProficiencyB1 0.11379  0.03810    2.986
```

Proficiency prediction model (baseline: Articulation rate + pause frequency)

```
Call:
multinom(formula = Proficiency ~ articulation_rate + pause_frequency, data = speakers)
Coefficients:

       (Intercept) articulation_rate pause_frequency
B2     4.431488    4.854783          -26.724735
C1     -37.569214  18.085977         -3.640931
native -36.969567  19.320178         -25.172595

Std. Errors:

       (Intercept) articulation_rate pause_frequency
B2     6.333065    3.333483          17.61253
C1     44.089429   12.520008         48.51376
native 44.924197   12.643535         51.43438

Residual Deviance: 38.97739
AIC: 56.97739
```

Proficiency prediction model (baseline + S_clause)

```
Call:
multinom(formula = Proficiency ~ articulation_rate + pause_frequency + Sclause, data = speakers)
Coefficients:

       (Intercept) articulation_rate pause_frequency Sclause
B2      -3.2375     20.90556          -77.06433       -13.64170
C1      -111.5640   58.95475          -53.02176       -23.32699
native -110.5393   60.15895          -76.28529       -23.93120

Std. Errors:

       (Intercept) articulation_rate pause_frequency Sclause
B2     14.53254    24.80746          77.31623        15.68490
C1     178.29325   53.95349          233.47834       18.96281
native 178.52429   53.98005          234.26668       19.02854

Residual Deviance: 26.18087
AIC: 50.18087
```

Proficiency prediction model (baseline + S_phrase)

```
Call:
multinom(formula = Proficiency ~ articulation_rate + pause_frequency + Sphrase, data = speakers)
Coefficients:

       (Intercept) articulation_rate pause_frequency Sphrase
B2      -1.935702   39.88477          -141.95300      -34.15078
C1      -173.583850 98.62193          -75.69711       -47.61274
native -169.577664 98.87231          -100.26668      -45.14241

Std. Errors:

       (Intercept) articulation_rate pause_frequency Sphrase
B2     23.03714    29.84157          95.52045        24.92308
C1     148.82571   62.36998          205.92816       31.86623
native 147.70944   62.19217          205.53023       31.81918

Residual Deviance: 21.41145
AIC: 45.41145
```

Proficiency prediction model (baseline + SPR)

```
Call:
multinom(formula = Proficiency ~ articulation_rate + pause_frequency + SPR, data = speakers)
```

```
Coefficients:

       (Intercept) articulation_rate pause_frequency SPR
B2     1.272329    7.551009          -37.95681       -6.774431
C1     -82.579342  36.205388         -32.34074       -28.704128
native -73.940424  37.894293         -87.56004       -8.367411

Std. Errors:

       (Intercept) articulation_rate pause_frequency SPR
B2     6.38891     5.556608          25.67418        5.356414
C1     89.44987    28.519258         78.54089        19.543573
native 88.61562    28.329395         90.41205        19.134584

Residual Deviance: 17.436
```

Address for correspondence

Sylvain Coulange
Grenoble Computer Science Laboratory, CNRS, Institute of Engineering
Université Grenoble Alpes
38000 Grenoble
France

sylvain.coulange@univ-grenoble-alpes.fr
https://orcid.org/0000-0002-9728-1181

Co-author information

Nivja H. de Jong
Leiden University Centre for Linguistics (LUCL)
Leiden University Graduate School of Teaching (ICLON)
Leiden University

n.h.de.jong@hum.leidenuniv.nl
https://orcid.org/0000-0002-3680-3820

Publication history

Date received: 29 July 2025
Date accepted: 3 February 2026
Published online: 13 March 2026
Corrected: 30 March 2026

In the original Online-First version of this article published on 13 March 2026, there was an error in Figure 2. These have been updated in the current version of the article.