The Temporal Structure of Intonation Units in English and Japanese: Covid-19 Conversations*

SHOICHI IWASAKI
University of California, Los Angeles

1 Introduction

Speakers produce speech in chunks. These chunks are identified as intonation units (IUs) in the functional linguistics tradition. IUs are identifiable by various prosodic cues such as an overarching intonation contour and surrounding pauses, and interface with grammar, cognition, and social interaction (Chafe 1979, 1980, 1988, 1994, 2018, Pawley & Syder 1983, Iwasaki 1993, Iwasaki & Tao 1993, Durie 1994, Croft 1995, 2007, Park 2002, Matsumoto 2003, Barth-Weingarten 2016, among others).[1] Past research has explored IUs from various perspectives, with three main areas receiving the most rigorous attention. The first area concerns the grammatical identification of IUs. IUs are often regarded as coextensive with a complete simple clause, a tendency described as the 'one clause at a time' principle (Pawley & Syder 1983, Chafe 1994). This principle suggests that a speaker can encode up to one simple clause in a single IU but not more. Exceptions exist for

[1] Units similar to IUs have been discussed by different names by different authors, such as 'tone units' (Crystal 1980) and 'tone groups' (Halliday, 1967). Phoneticians/phonologists are also interested in a similar unit. Selkirk (1984) uses the term 'intonational phrase', while Cruttenden (1997) uses 'intonation group', and those who work under the ToBI system use the terms 'intonation phrase' and 'intermediate phrase' (Beckman, Hirschberg, & Shattuck-Hufnagel 2005).

* I would like to acknowledge Caleb Williams and An Nguyen for their assistance in preparing this paper. I take full responsibility for any remaining errors in this paper.

Japanese/Korean Linguistics 31.
Edited by William Giang, Lucien Brown, Shimako Iwasaki, Satoshi Nambu, and Daniel Pieper.
Copyright © 2024, Shoichi Iwasaki.

formulaic expressions containing longer or more complex structures. Givón (1984: 258–63), on the other hand, proposed the 'one chunk per clause' principle, which states that only one conceptual chunk (e.g. a noun, a verbal, or adjectival predicate, or an adverbial phrase) can be asserted per clause. This led Chafe (1994: 108–19, 2018: 62) to focus on a second area of interest in IU study, referred to as the 'one new idea' constraint. When multiple new ideas are introduced, a single clause is often divided into multiple IUs. The third key area of research concerns the physical length of an IU, which has been estimated in terms of both syllable count and temporal duration. In this area of inquiry, both the length of IUs and pre-IU pauses were examined. On average, an IU is estimated to last 'no more than a second or two' (Chafe 2018: 58), while the pre-IU pause is minimally around 0.2 seconds. However, research validating these estimates across languages beyond English remains limited.

This study aims to address these gaps by comparing IU length and pre-IU pause duration in English and Japanese. The analysis reveals that Japanese IUs are longer than their English counterparts in both IU duration and pre-IU pause length. This difference is attributed to morphological and pragmatic factors. Japanese relies heavily on multisyllabic words, whereas English favors monosyllabic ones, inherently leading to longer IUs in Japanese. Furthermore, Japanese clauses tend to encode greater pragmatic and interactional information, which not only extends IU duration but also increases pre-IU pause length due to the additional cognitive processing required.

The paper is structured as follows. Section 2 describes the database and methodology used in this study. Section 3 reviews previous research on IU length and pre-IU pause duration. Section 4 presents the findings and discusses their significance. Finally, Section 5 summarizes the study and offers concluding remarks.

2 Data and Methodology

For this study, I use a subset of the Japanese and English data from the Multilingual Covid-19 Conversation Data Corpus (hereafter referred to as the Covid-19 Data).[2] This corpus consists of semi-controlled natural conversations in multiple languages and is well suited for examining grammatical, cognitive, and interactional aspects of IUs from a crosslinguistic perspective. The dataset includes thirty conversations in each of six languages: English, Japanese, Korean, Mandarin, Cantonese, and Thai. In each recording, two unacquainted native speakers, aged between 19 and 35, were invited to converse for twenty minutes on Zoom, focusing on the topic of the Covid-19 pandemic. Most conversations were recorded between July 2021 and March 2023, during a period when the pandemic was gradually coming under control.

This dataset has three key features: First, it represents a form of 'disaster discourse', in which speakers share personal experiences and express individual views about the pandemic as a disaster. Because participants are typically eager to talk about their experiences, the conversations tend to be spontaneous and natural. Second, the data capture 'first encounter discourse', where participants do not share personal common ground but instead build rapport based on a perceived communal common ground (Clark 1996). This provides researchers with valuable insight into how unshared

[2] The Multilingual Covid-19 Conversation Data Corpus Project was cosponsored by UCLA's Academic Senate and the Terasaki Center for Japanese Studies, with Shoichi Iwasaki as PI and eighteen collaborating members. It is also cosponsored by 'Grant-in-Aid for Scientific Research (B) 2022-2024' with Kazuyo Murata, Ryukoku University (Japan), as PI and four other researchers.

information is introduced and negotiated in interaction. Finally, the dataset offers a collection of multilingual conversations gathered under uniform conditions, enabling systematic crosslinguistic and crosscultural comparisons. Previous IU research has drawn from diverse sources, including narratives, informal conversations, public and personal telephone calls, and task-based dialogues, often involving friends, family members, or strangers. However, for meaningful crosslinguistic comparison, data comparability is crucial. The Covid-19 Data achieve this by sharing consistent situational characteristics (cf. Biber & Conrad 2019), including medium (Zoom), participant relationship (strangers), setting (virtual), and topic (the Covid-19 pandemic).

Each conversation was manually transcribed, with some initially processed using an automatic transcription tool such as pyTranscriber. These initial transcriptions were then retranscribed based on Du Bois's transcription system (Du Bois 1991, Du Bois et al. 1992) by a trained annotator and subsequently verified by a second transcriber for accuracy. The final transcripts were produced after measuring the durations of pre-IU pauses and the lengths of IUs using Audacity.[3] In some cases, IU identification was verified using PRAAT.[4] A list of key transcription symbols is provided in Appendix (A).

IUs are usually marked by audible pauses of more than 0.2 seconds, but even in the absence of clear pauses, they can be identified through a range of prosodic and verbal cues. These include a coherent intonation contour (melody); boundary tones (tail pitch movements; Nagahara & Iwasaki 1994); pause fillers; pitch resets; anacrusis (acceleration of unstressed, IU-initial elements); final vowel lengthening; articulation speed (tempo); voice quality; and disfluency behaviors, such as false starts and repetitions. In the transcripts, IU boundaries indicated by cues other than pauses are denoted by (0). Boundaries coinciding with a micro pause of less than 0.2 seconds are marked with two dots (..), while those exceeding 0.2 seconds are represented by their exact duration in parentheses (e.g. (0.2), (2.5)).[5]

For this study, I analyzed three English conversations (ECO#001 [f-m], ECO#003 [f-f], ECO#004 [f-f], each 20 minutes) and four Japanese conversations (JCO#001 [f-m, 20 minutes], JCO#016 [f-m, 20 minutes], JCO#017 [f-f, 10 minutes], JCO#022 [m-m, 10 minutes]). From these conversations, a total of 3,084 English IUs and 2,046 Japanese IUs were extracted. These were categorized into three main types: Substantive IUs (either clausal or phrasal); Regulatory IUs; and Nongrammatical IUs. Table 1 presents the distribution of these IU types in both English and Japanese.

[3] See https://www.audacityteam.org/.

[4] See https://www.fon.hum.uva.nl/praat/.

[5] A useful framework for identifying IUs is provided in Himmelmann et al. (2018). While they use the term 'intonational phrase' (IP), this paper continues to use IU for consistency. According to Himmelmann et al. (2018: 213), IUs are 'distinct units perceivable by means of a coherent melody'. The perception of such a melody can be disrupted by temporal cues (e.g. pauses, final lengthening) or pitch cues (e.g. pitch resets), which may operate interdependently—sometimes one cue overriding the other. For instance, a brief pause may not break an IU if a strong melodic contour bridges it; conversely, a long pause (over one second) typically prevents melodic coherence. When multiple cues converge, the boundary becomes more perceptually salient. Much like a Gestalt figure, IU boundaries may vary based on which cues a listener focuses on. However, in most cases, a dominant perceptual figure emerges due to the convergence of break-indicating features. See Barth-Weingarten (2016) for an extended review of issues surrounding IU identification.

	English	Japanese
Substantive	1531 (49.6%)	1444 (70.6%)
Clausal	*(806 (26.1%))*	*(703 (34.4%))*
Phrasal	*(725 (23.5%))*	*(741 (36.2%))*
Regulatory	847 (27.5%)	502 (24.5%)
Nongrammatical	706 (22.9%)	100 (4.9%)
TOTAL	3084 (100 %)	2046 (100%)

Table 1: The Distribution of Different Types of IUs in English and Japanese

Substantive IUs are those that convey information about events, states, and referents. Clausal Substantive IUs are divided into three subtypes: Full clausal IUs, Subjectless clausal IUs, and Complex clausal IUs. A few examples of each from English and Japanese are listed below. The number or the two-dot notation within parentheses at the beginning of each line represents the pre-IU pause length, while the number in the square brackets at the end denotes the length of IUs. A list of abbreviations used for glosses is found in Appendix (B).

1. Full clausal IUs consist of both a subject and a predicate.

(..) I don't take very good care of myself? [1.5]
(..) he's a balloon artist. [1.1]
(0.5) kinoo watashi tomodachi to asonda-n-desu kedo: ^ [2.1]
yesterday 1SG friend with play-PAST-SE-COP but
'I went out with my friend yesterday.'
(0.2) iran mono ippai dete-kimashita. [1.7]
useless thing many come-out-PAST
'Many useless things appeared.'

2. Subjectless clausal IUs lack a subject but still express a clausal idea.

(..) to be in a pandemic? [0.9]
(..) watch movies and stuff like that, [1.2]
(0.9) nanka (..) kengai no daigaku na node (..) ano, [2.2]
INJ out-of-prefecture GEN university COP so INJ
'uhm, (my university) is out of the Prefecture.
(1.0) tsu tsu ni arimasu. [0.8]
(city name) (city name) LOC exist:POL
'It's in the City of Tsu.'

3. Complex clausal IUs include coordinate or complement clauses.

(..) overall I think I'm (..) been pretty lucky like, [2.1]
(..) I'm really thankful that I had him there ? [2.1]

(1.1) nagoya to wa chigau to omoimasu. [1.4]
(city name) with TOP different QT think:POL
'I think it is different from Nagoya.'
(..) tomodachi toka tsukuri nikui janai desu ka. [2.2]
friend etc. make difficult NEG COP Q
'Making friends is difficult, right?'

Phrasal IUs include noun phrases (NPs) with or without prepositions (for English); postpositions (for Japanese); or modifiers, as well as other types of standalone phrases and words. Regulatory IUs serve functions such as encoding textual ('well'), interactional ('you know'), cognitive ('let me see'), or validational ('maybe') information (Chafe 1994: 63). Some IUs in conversational data do not constitute a grammatical unit—for example, truncated utterances or ungoverned combinations of fragments. These are coded as 'Nongrammatical' in Table 1.

The next is an excerpt that contains Subjectless Clausal (SLC), Phrasal (PHR), Nongrammatical (NGR), and Regulatory (REG) IUs.

(1) ECO#001

991	M1:	(1.0)	that was-	[0.4] NGR
992	M1:	(..)	really,	[0.5] PHR
993	M1:	(..)	I guess,	[0.3] REG
994	M1:	(..)	interesting if I sort of ,	[1.1] NGR
995	M1:	(0.4)	dissociate myself from it.	[1.0] SLC

Chafe (1994: 66) reports that in his data, 'the mean proportion of single-clause substantive intonation units … is about 60%'. Croft (1995: 845), in a more detailed quantitative analysis of 1,989 IUs from Pear Story data (see Chafe 1980)[6], reported that 53.7% were single-clause Substantive IUs (see also Croft 2007: 12).[7] As shown in Table 1, clausal and phrasal Substantive IUs make up the majority of IUs in the current data for both languages—49.6% in English, and 70.6% in Japanese. The ratio of clausal to phrasal IUs is also comparable: 26.1% versus 23.5% in English, and 34.4% versus 36.2% in Japanese. Regulatory IUs also appear at similar rates—27.5% in English, and 24.5% in Japanese. One striking difference is the frequency of Nongrammatical IUs, which account for 22.9% of the English data but only 4.9% of the Japanese data.[8] For the purposes

[6] The *Pear Story* is a six-minute film depicting a man harvesting pears from a tree while a boy secretly steals some of them. The film also includes several interactions involving other participants. Though it contains ambient sounds, it features no speech. Chafe and his colleagues created the *Pear Story* film in 1975 as a tool for eliciting narratives from speakers of different languages for discourse and linguistic analysis.

[7] This includes 'simple clauses with subjects, subjectless simple clauses, relative clauses, adverbial clauses' but does not include 'complex and coordinate clauses'.

[8] Croft (1995: 844) observes that only 3% of the IUs in his English dataset are classified as nongrammatical units (non-GUs). These encompass false starts, incomplete constituents, and disjointed IUs—defined as IUs composed of two elements that do not collectively form a coherent grammatical unit. The higher proportion of nongrammatical IUs in the present study is likely attributable to the nature of the data. While Croft's (1995) analysis focused on narrative monologues (specifically the Pear Story), the dataset analyzed in this study is derived from spontaneous, interactive conversation, which inherently exhibits a greater frequency of disfluencies and interruptions. The reason for the lower

of this study, only Substantive IUs were included in the analysis, yielding a total of 1,531 English IUs and 1,444 Japanese IUs as the target dataset.

3 Temporal Structure of Intonation Units

The temporal aspect of IUs involves two components: the duration of an IU itself, and the length of the pause preceding it. Each will be discussed in turn below.

3.1 Length of Intonation Units

Pawley and Syder (1983: 202) report that IUs (or what they term 'one fluent clause') typically consist of four to ten words. Chafe (1994: 64–5) offers a similar finding, reporting an average of 4.84 words per English Substantive IU.[9] In a later study, Chafe (2018) measured IU length in temporal terms and concluded that IUs 'occupy no more than a second or two' (ibid.: 58), with an average duration of approximately 1.5 seconds (ibid.: 66). This aligns closely with the mean IU length of 1.680 seconds (with a median of 1.227 seconds) found in the analysis of 50,510 interpausal units (IPUs) in the Switchboard corpus (Levinson & Torreira 2015: 16). IPUs, defined by pauses of 0.18 seconds or longer, are functionally similar to IUs. Levinson (2016: 6) also notes that 'short phrase- or clause-like units' (i.e. IUs) used in interaction tend to be about 2 seconds in length.

Other studies have explored similar speech units using different conceptual frameworks. For instance, Enfield (2025) introduces the 'enchronic envelope', a unit defined within interactional time that captures the temporal window for social action through language. He suggests that this unit, often composed of a single clause, is comparable to the IU (ibid.: 7). Drawing from cognitive science research (e.g. Pöppel 2009), he further proposes that the typical duration of an 'enchronic envelope' is approximately 2.5 seconds. He equates this temporal window to the 'phenomenal present', which reflects the subjective experience of the present moment. This view resonates with Chafe's (1994: 63, 2018: 58) concept of 'foci of consciousness', though Enfield's (2025) estimate is longer than Chafe's 1.5-second average or the Switchboard corpus average of 1.680 seconds just mentioned. More recently, Inbar et al. (2020: 15847) investigated IU duration across six typologically diverse languages,[10] and found the median IU length in English to be 0.94 seconds, while other languages—except Hebrew—clustered around a median duration of approximately 1 second. Hebrew exhibited the shortest median IU length at 0.58 seconds.

In sum, existing research suggests that the typical median duration of English IUs falls between about 1 and 2.5 seconds, depending on the type of data and measurement approach. Based on the data analyzed in this study, I will argue in Section 4 that English IU length tends toward the lower end of this range while Japanese IU length is slightly, but statistically significantly, longer than that.

---

frequency of Nongrammatical IUs in the Japanese data is deferred for future discussion; however, one plausible explanation is that Japanese speakers, particularly in conversational contexts involving unfamiliar partners, tend to exhibit a more deliberate speech style compared to their English-speaking counterparts. This caution exercised by Japanese speakers may result in fewer instances of Nongrammatical IUs.

[9] It is 1.36 words for English Regulatory IUs.

[10] Inbar et al.'s (2020) study used a mixture of natural conversation and spontaneous narrative in English, Hebrew, Russian, Papuan Malay, Wooi (Austronesian), and Yari (Papuan).

3.2 Pre-IU Pauses

Researchers have long been interested in the length of pre-IU pauses (e.g. Goldman-Eisler 1967), which can occur within a speaker's turn (intraspeaker pauses) or between speakers (interspeaker pauses). In narrative discourse, two types of intraspeaker pre-IU pauses are commonly distinguished: external and internal pauses (Himmelmann et al. 2018: 213). External pauses, also known as planning pauses, occur between two adjacent IUs, giving the speaker time to prepare the upcoming utterance. Internal pauses, or hesitation pauses, on the other hand, occur after an IU is launched and often cooccur with word searches and self-repair behaviors, reflecting cognitive or social difficulty. Importantly, an internal/hesitation pause may transition into an external/planning pause if the speaker abandons the current IU.

Chafe (1979) observed that the length of an external pause correlates with the cognitive load required to produce the following IU. For example, in the Pear Story narrative, one speaker begins to describe the pear film with a long initial pause:

(2) (4.25) Um .. it starts out .. there's a (3.3) well, (1.45) the—landscape is like …
(Chafe 1979: 167, 1980: 310)

The first pause in (2) (4.25 seconds) is an external pause, reflecting the speaker's effort to organize her thoughts before commencing the story. The second pause (3.3 seconds), initially a hesitation pause, evolved into a planning pause, since her effort to find the right descriptive term (e.g. 'a farmland', 'an orchard') failed, as surrounding elements such as false starts ('there's a') and hesitation markers ('well') reveal. Long external pauses also appear at episode boundaries in narratives when a scene with new characters begins. The speaker explicitly comments that she is switching to a new scene after a long pause both in (3) and (4).

(3) And then (5.2) tsk so—.. then we switch to the boy riding on the bicycle (Chafe 1980: 177).
(4) (0.9) And – (2.9) the next people .. who come by (ibid.: 179)

In contrast, shorter pauses—typically between 0.2 and 0.6 seconds—occur between IUs that share a 'coherent center of interest' (Chafe 1994: 140).[11] These patterns in narrative monologues can be contrasted with turn-taking behavior in conversations. Unlike narrators, participants in interactive conversations tend to rely on shorter pauses to avoid losing their turn. (Cases of longer pauses preventing speaker change will be discussed in Section 4.2.2.)

In Conversation Analysis (CA), researchers are particularly interested in interspeaker pauses or transition pauses. Enfield (2017: 38–40) cites studies of telephone conversations in Dutch (de Ruiter et al. 2006), English (Levinson & Torreira 2015), and German (Riest et al. 2015), showing that the average offset time at the speaker change is around 200 milliseconds.[12] Although pauses can be longer, 1 second is typically regarded as the 'standard maximum silence' (Jefferson 1989,

[11] The 'center of interest' is a cognitive phenomenon that corresponds to a segment of discourse composed of multiple IUs, which cohesively adhere to a shared thematic framework.

[12] The offset time includes both overlap (a negative offset) and a gap (a positive offset) between the end of one speaker's utterance and the beginning of the other speaker's utterance.

Enfield 2017: 70–1).[13] Stivers et al. (2009: 10588) studied ten languages, including English and Japanese,[14] and found a mean response time of 207 milliseconds following yes/no questions. This is consistent with findings from Heldner and Edlund (2010), who report similar average turn-transition times in corpora from Dutch, Swedish, and English (see also Levinson & Torreira 2015: 15).

3.3 Summary and Prospect

Previous studies have found that the average length of an IU is approximately 1 second, with an upper limit of around 2.5 seconds. Similarly, the average duration of a pre-IU pause is minimally about 0.2 seconds, with an upper threshold of 1 second. While many recent studies have investigated IU and pre-IU pause lengths across multiple languages to establish crosslinguistic consistency, this paper offers a complementary perspective by focusing on a comparative analysis of IU and pre-IU pause durations in English and Japanese. The aim is not only to demonstrate that the observed patterns in both languages fall within the universal range identified by previous research but also to highlight that subtle, yet statistically significant, differences exist. I argue that these differences stem from structural and pragmatic distinctions between the two languages and therefore, warrant further examination within the broader framework of language-specific interactional practices.

4 Results and Discussion

4.1 IU Length

In this section, I report the results of the statistical test that compared the lengths of IUs in English and Japanese. A total of 1,531 IUs were extracted for English and 1,444 for Japanese for general analysis (see Table 1 in Section 2). However, 125 IUs from the English data and 52 IUs from the Japanese data were excluded because their lengths could not be measured accurately due to overlapping speech, accompanying laughter, or other distracting factors. Consequently, the remaining 1,406 English IUs and 1,392 Japanese IUs were included in the comparison. Table 2 presents the results.

[13] Interspeaker pauses may have various social meanings. Templeton et al. (2022) show that a shorter gap indicates a social connection between acquainted interlocutors while a long pause either indicates awkwardness between strangers or provides a space for enjoyment among friends.

[14] Other languages included in the study are Lao, Korean, Italian, Danish, Dutch, Yélî Dnye (Papua New Guinea), Tzeltal (highland Mexico), and ǂĀkhoe Hailom (Nambia). Although, on average, speakers across the ten languages take 207 milliseconds to respond, Japanese speakers respond in just 7 milliseconds. The underlying causes of this exceptional brevity in the turn-transition gap among Japanese speakers warrants further investigation, which I will defer to a future research agenda. I extend my gratitude to Makoto Hayashi for the insightful discussions on this topic during and after JK 31.

Measure	English (sec.)	Japanese (sec.)
Mean[a]	1.11	1.37
Median[b]	1.00	1.30
Mode	0.90	1.00
Number of IUs	1406	1392

[a] Mean difference statistically significant (Welch Two Sample t-test, $p < 0.05$)
[b] Median difference statistically significant (Bootstrapped 95% CI does not include (0))

Table 2: Lengths of English and Japanese IUs

According to Table 2, the median length of English IUs is 1 second, closely aligning with the 0.94-second median reported by Inbar et al. (2020: 15847) for English IUs (see Section 3.1). In a separate study, Inbar et al. (2023) suggest that the neural system is attuned to a rhythm of one IU per second, based on EEG responses from listeners who heard a short narrative in their native language, Hebrew.

Table 2 also indicates that the median length of Japanese IUs, 1.30 seconds, is statistically significantly longer than that of English IUs, 1 second (see Appendix (C) for statistical results). What accounts for this difference? I propose that it stems from morphological differences between the two languages. Japanese words are generally longer than English words, particularly among high frequency words. High frequency words in English are often monosyllabic, while their Japanese counterparts tend to be multisyllabic. To demonstrate this pattern, I compared the major lexical categories—nouns, verbs, adjectives, and adverbs[15]—among the 100 most frequent words in two datasets: ECO#001 (English) and JCO#001 (Japanese). Tables 3 and 4 present the type frequencies of these words by syllable length. In total, 50 types were identified in English and 29 in Japanese (see Appendix (D) for the full word lists).

English	1 syllable	2 syllables	3 syllables	4 syllables	TOTAL
Nouns	19	2	2	0	23
Verbs	14	2	0	0	16
Adjectives	1	1	0	0	2
Adverbs	5	3	0	1	9
Total	39 (78%)	8 (16%)	2 (4%)	1 (2%)	50 (100%)

Table 3: Distribution of Words According to Syllable Count in English

Japanese	1 syllable	2 syllables	3 syllables	4 -5 syllables	TOTAL
Nouns	0	5	5	2	12
Verbs	0	1	0	3	4
Adjectives	0	0	0	0	0
Adverbs	0	5	5	3	13
Total	0 (0%)	11 (38%)	10 (34%)	8 (28%)	29 (100%)

Table 4: Distribution of Words According to Syllable Count in Japanese

[15] Other words are minor words such as prepositions in English and case particles in Japanese. They also include many interjections, such as *like* in English and *nanka* in Japanese. See Footnote 18.

These two tables clearly highlight the monosyllabic tendency among frequent major lexical categories in English, in contrast to Japanese. Figure 1 compares the overall distribution of syllable counts in the two languages, based on the percentage of total words in the rows of Tables 3 and 4. In English, 39 words (78%) were monosyllabic, while 11 words (22%) were multisyllabic. In contrast, Japanese had no monosyllabic words (0%), with all 29 words (100%) being multisyllabic.

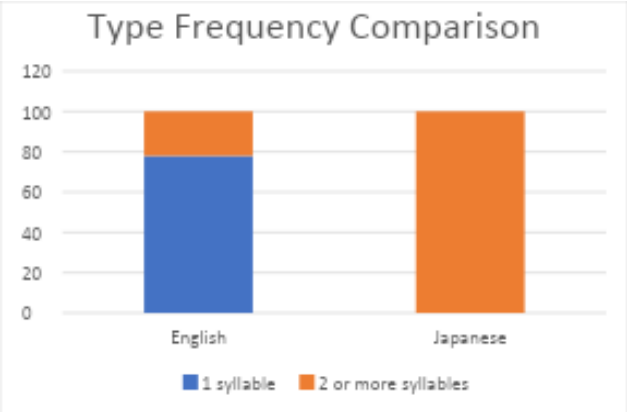

Figure 1: Type Frequencies of Monosyllabic and Multisyllabic Words Among Major Lexical Categories in the Top 100 Words from ECO#001 and JCO#001

The difference becomes even more striking when we compare the total token counts of monosyllabic words as opposed to multisyllabic words, as Table 5 and Figure 2 show.

	1 syllable	2+ syllables	Total
English	707 (85.3%)	122 (14.7%)	829 (100%)
Japanese	0 (0%)	76 (100%)	76 (100%)

Table 5: Token Frequencies of Monosyllabic and Multisyllabic Words Among Major Lexical Categories in the Top 100 Words from ECO#001 and JCO#001

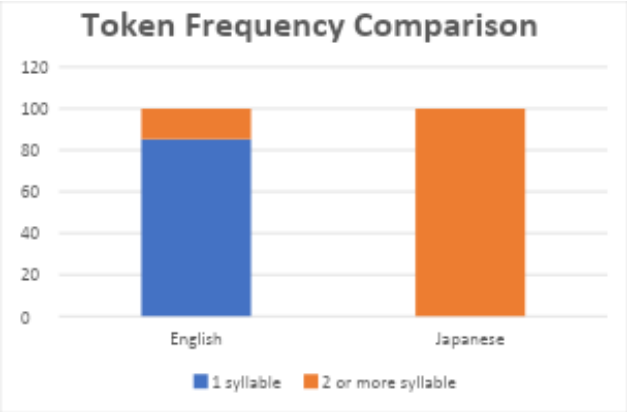

Figure 2: Token Frequencies of Monosyllabic and Multisyllabic Words Among Major Lexical Categories in the Top 100 Words from ECO#001 and JCO#001

While 707 high frequency words (85.3%) in English were monosyllabic, no such monosyllabic words were identified in Japanese. Consequently, to convey the same amount of information, Japanese generally requires more time to accommodate the additional syllables,[16] assuming the production time per syllable is approximately equivalent.[17] For instance, consider the sentences 'I like eels' in English and *(watashi wa) unagi-ga suki-da* in Japanese. The English sentence comprises three syllables, whereas its Japanese counterpart contains seven syllables, even after excluding the subject through ellipsis. To illustrate this in a realistic context, observe a pair of similar IUs from the data. Both IUs commenced after a 0.1-second pause. The number in square brackets following the IU line indicates its duration in seconds, followed by the number of syllables in the IU. The English IU has six syllables and is 1 second in length (the exact median length in Table 2), while the Japanese IU has nine syllables and is 1.4 seconds (0.1 seconds longer than the median value in Table 2).

(5)	ECO001 1301 F1:	(0.1) the pandemic happened,	[1] (6 syllables)
(6)	JCO#001 142 F1:	(0.1) korona-ka natte shimatte^	[1.4] (9 syllables)
		Covid-pandemic become-ASP	
		'The Covid pandemic started.'	

In summary, I propose that Japanese IUs are longer than English IUs by 0.3 seconds in general due to the necessity of accommodating multisyllabic words in Japanese, whereas English predominantly utilizes monosyllabic words.

4.2 Pre-IU Pause Length

In addition to IU length, the two languages also differ in pre-IU pause length. A pre-IU pause refers to the silent gap before an IU begins. Some IUs naturally lack this gap when they are identified based on nonpause features such as pitch reset, acceleration of syllables, or changes in volume (see Section 2). These IUs, indicated by the notation (0) in the transcript, were excluded from the statistical analysis. Specifically, 553 IUs in the English data and 103 in the Japanese data were removed, leaving 978 and 1,341 IUs, respectively, in the two languages. Additionally, micro pauses transcribed with two dots (..) were standardized to (0.1) for the statistical test. Table 6 presents the results.

[16] Pellegrino et al. (2011) examined this phenomenon from the opposite perspective, arguing that when information density (ID)—defined as the amount of linguistic information encoded per syllable—is low, speech rate (SR) must consequently be high (see also Coupé et al. 2019). To convey an equivalent amount of information within the same time frame (e.g. one second), Japanese speakers must produce more syllables, that is, they must speak at a faster rate than Mandarin and English speakers. However, their study was based on read-aloud translations of English texts across multiple languages, including Japanese. As a result, their findings are not directly comparable to those of the present study, which examine naturally occurring conversational data.

[17] According to Levinson (2016: 7), the average duration of a single syllable is 200 milliseconds across languages.

Measure	English (sec)	Japanese (sec)
Mean[a]	0.28	0.42
Median[b]	0.10	0.30
Mode	0.10	0.10
Number of pre-IU pauses	978	1341

[a] Mean difference statistically significant (Welch Two Sample t-test, $p < 0.05$)
[b] Median difference statistically significant (Bootstrapped 95% CI does not include (0))

Table 6: Lengths of Pre-IU Pauses in English and Japanese

4.2.1 Pre-IU Pause for IU Launch Preparation

According to Table 6, the mean and median length of a pause preceding a Substantive IU is longer in Japanese than in English. The difference is statistically significant (see Appendix (C)). One reason for the shorter pauses in English can be attributed to the presence of 'clause initiators' in the language. Clause initiators include conjunctions (e.g. 'and', 'so', 'cuz'), discourse markers (e.g. 'you know'), and hesitation markers (e.g. 'uh=m'). These elements compensate for shorter pauses following an intraspeaker pause. Speakers can extend the preparation period by combining the duration of a pause with a clause initiator to prepare the forthcoming utterance. Observe the examples of clausal IUs from ECO#001, where clause initiators are capitalized and underlined for easy recognition.

(7) ECO#001

(..)	AND he was a performer ,	[1.3]
(..)	AND I live by myself ?	[0.9]
(..)	SO I was like at the dance studio ,	[1.5]
(..)	CUZ it had lasted for so lo=ng.	[1.6]
(..)	CUZ I know so many people ,	[1.0]
(..)	OF COURSE the dance studios closed.	[1.7]
(0.3)	AND it wasn't any different.	[0.9]
(0.5)	UH=M (..) I still like teach with chalkboard ,	[2.1]
(0.6)	AND I'm an introvert ?	[0.9]

Japanese speakers use clause final conjunctive particles (e.g. *kedo*, *kara*) and conjunctive forms (e.g. the *-te* form) instead of clause-initial conjunctions. Thus, they cannot use the same strategy as English speakers.[18]

[18] However, Japanese speakers frequently employ the interjection *nanka* at the beginning of an IU. Similarly, English speakers regularly use the comparable interjection *like*. The use of *nanka* is extremely frequent, with a total of 236 tokens across the four Japanese datasets, occurring not only in clause-initial positions but also in other positions within an IU. This contrasts with other interjection such as *ano* (60 tokens) and *eeto* (3 tokens). Similarly, the use of *like* is very common among English speakers, with 858 instances found in various positions within the three English datasets. The relationship between gap length and the use of clause initiators warrants further investigation.

There is another reason why pre-IU pauses in English are shorter than those in Japanese. Chafe (1979) suggests that pauses facilitate cognitive work in preparing IU production. Thus, the data in Table 6 imply that the longer pauses in Japanese reflect a more complex mapping of information onto syntax. Compare the examples in (8) and (9).

(8) ECO#001

(..)	I'm in my apartment (..) a lot.	[1.5]
(..)	how's your pandemic been.	[1.0]
(..)	he's a balloon artist.	[1.1]
(..)	I don't take very good care of myself?	[1.5]
(..)	I live off pickles and coffee?	[1.4]

(9) JCO#001

(0.2)	masuku wa utte-nakatta **desu yone.** mask TOP sell-NEG-PAST COP PP 'They were not selling masks, right?'	[1.4]
(0.8)	mise shima**ttari shite,** store close-so.on do:TE 'Stores also closed etcetera, and …'	[1.2]
(0.3)	nikibi dekitari **toka:^** pimples come-out so-on '(I) got pimples, and ….'	[1.4]
(0.5)	kinoo watashi tomodachi to asonda-n-desu **kedo: ^** yesterday 1SG friend with play-PAST-SE-COP but 'I went out with my friend yesterday, but.'	[2.1]

Notice that the clauses in (8) primarily represent barebone propositional information. In contrast, the clauses in (9) encode extrapropositional information through sentence-final elements, such as *desu yone* (politeness and strong assertion); *tari shite* (hesitation and continuation); *toka* (softened assertion); *kedo* (softened assertion); as well as *masu* (politeness); *desu* (politeness); *ne* (confirmation); and *wake* (cohesion) (see Tanaka 1999: 85, 92). Speakers of Japanese must take into account sociolinguistic factors (e.g. honorifics and speech levels), pragmatic and discourse information (e.g. topicalization and focalization), as well as case relations among clause elements when preparing for an upcoming IU. While English speakers may also consider such extrapropositional information, their commitment appears less evident according to the examples in (8).

If speakers fail to prepare a complete IU during the pre-IU pause, they must modify the IU in real time. The next excerpt illustrates what needs to be planned during the pause.

(10)

179	F1	(0.7)	(H) ma sakini ano hataraiteru:: ? INJ before INJ work:ASP 'those who have been working	[1.8]

180	F1	(0.8)	kata ni ^ person to/from 'to/from the person'	[0.4]
181	F1	(..)	kara kiite:: from hear:TE 'from the person, (I) heard.'	[0.7]

Translation:
179 F1: (0.7) (H) And, (I hear from a senior coworker who) has been working longer than I have.
180 F1: (0.8) from/to that person HONORIFIC
181 F1: (.) heard it FROM her

The speaker took a 0.7-second pause accompanied by an inhaling noise before initiating the first IU, which concludes with a modifying verb (*hataraiteru* 'has been working') without a modified head noun (*kata* 'person').[19] The break between a modifier and its head is atypical, but the speaker paused for an additional 0.8 seconds before producing the head noun. This long pause can be attributed to a sociolinguistic concern regarding the choice between the nonhonorific form, *hito* 'person', and the honorific form, *kata* 'person', for the head noun. Furthermore, the speaker revised the case frame by replacing the particle *ni* with *kara* 'from'.[20] If the utterance had been perfectly planned (as in a written discourse) during the pre-IU pause before the first IU, it could have been produced as follows.

(11) sakini hataraiteiru kata kara kiite
before work:ASP person HONORIFIC from hear:TE
'I heard (such anecdotes) from my senior co-worker.

However, the speaker is also constrained by the limit of IU length and cannot extend the duration significantly beyond a certain point (i.e. 1.30-second median length; see Table 2 in Section 4.1).

4.2.2 Pre-IU Pause Length and Speaker Change

In addition to its function for planning an upcoming clause, a longer pause can also serve an interactional purpose. Refer to Table 6 again (repeated below). Although short pauses (0.1-second modal length) are frequent in both languages, Japanese includes longer pauses than English, as evidenced by the median and mean lengths of pauses in the two languages. I will now examine the mechanisms of conversation to evaluate the significance of longer pauses in Japanese.

[19] The rising tail pitch movement on this IU (line 179) is significant. Ono et al. (1998) use the term 'half question' to describe this type of utterance delivery, where a speaker suspends an utterance at a nongrammatical boundary.
[20] Particle *ni* is ambiguous. It may mean 'to (the person)' or 'from (the person).' In this context, the speaker had heard some anecdote *from* a senior worker. *Kara* clarifies this meaning.

Measure	English (sec)	Japanese (sec)
Mean[a]	0.28	0.42
Median[b]	0.10	0.30
Mode	0.10	0.10
Number of pre-IU pauses	978	1341

[a] Mean difference statistically significant (Welch Two Sample t-test, $p < 0.05$)
[b] Median difference statistically significant (Bootstrapped 95% CI does not include (0))

Table 6: Lengths of Pre-IU Pauses in English and Japanese

Longer pauses are likely to trigger speaker changes. As noted in passing in Section 3.2, a one-second silence is the maximum length people can tolerate when waiting for a response to a question (Jefferson 1989). This implies that pre-IU pauses exceeding one second will strongly elicit a response from the other participant. Consequently, frequent speaker changes are expected beyond this threshold. To confirm this, I examined Japanese data (JCO#016) to investigate the relationship between pause length and speaker shifts.

Pre-IU pause (sec)	Sp-change	No Sp-change	total
3 - 4.8	5 (62%)	3 (37.5%)	8
1	14 (66.6%)	7 (33.3%)	21
0.5	16 (35.5%)	29 (64.4%)	45
0.1	13 (6.5%)	184 (93.4%)	197

Table 7: Pause Length and Speaker Change in Japanese (JCO#016)

When the pause length reached the threshold of 1 second, speaker change occurred 66.6% of the time. In contrast, when the pause length was below the threshold, speaker change occurred only 35.5% of the time or less. Among the 8 instances of long pauses (3 to 4.8 seconds, as shown in Table 7), speaker change occurred five times (62%). In 3 cases, the addressee took a turn to expand the current topic or shift to another topic.

(12) Topic shift (JCO#016)

544	M	(1.7) pasokon mo netsu mochimasu shi ne. laptop also heat hold:POL also PP 'Laptops also heat up.'	[1.2]
545	M	(..) @@@	[0.4]
546	F	(..) <@ nee @> PP 'Right.'	[0.5]
547	F	(..) soo desu ne. so COP PP 'That is right.'	[0.7]
548	M	(3.9) **nanka senkoo tte nani yarareteru-n-desu ka** INJ major TOP what do:ASP-SE-COP PP 'Uhm, what are you majoring in?'	[2.0]

In two cases, the addressee responded to the current speaker after a prolonged silence.

(13) Delayed response (JCO#016)

633	M	(..)	meebo mitara nanka,	[1.1]
			roster see:COND INJ	
			'When I saw the roster,	
634	M	(..)	aa nanka hanbun,	[0.9]
			oh INJ half	
			'Oh, the half (of it)	
635	M	(..)	kotchi kara shita minna koo,	[0.9]
			here from down all INJ	
			'All (names) from here down'	
636	M	(..)	aa chuugoku-kee no ppoi kanji no onamae ya na to,	[2.8]
			oh Chinese-group GEN like feeling GEN name:HON COP PP QT	
			'(I thought,) "Oh, they are like Chinese like names."'	
637	F	**(3.2)**	**hee.**	[0.3]
			INJ	
			'Woow.'	

In three cases, no speaker change occurred, and the current speaker self-selected to continue or terminate the current topic after a long pause. This type of turn exchange leaves an impression of an unsuccessful interaction.

(14) Self-select topic closure (JCO#016)

477	M	(..)	nanka masuku shite-nai hito.	[1.1]
			INJ mask wear-NEG person	
			'People without a mask,	
478	M	(0.3)	gyakuni ooi na tte-yuu no wa.	[1.5]
			opposite many PP COM SE TOP	
			'on the contrary, are many,	
479	M	(..)	inshoo ari[masu ne.	[NA]
			impression exist PP	
			'I have such an impression.'	
480	F		[arimasu yone.	[NA]
			exist PP	
			'I think so, too.'	
481	M	(0.5)	n:::	[0.4]
			INJ	
			'Hmm'	
482	M	**(3.2)**	**masuku ka.**	[0.9]
			mask Q	
			'Masks…'	

Overall, long pre-IU pauses in the current Japanese data appeared within the context of the turn-taking system, contrasting with those that appear in narrative contexts for IU preparation. In contrast, longer pauses for English serve functions beyond the turn-taking system. Earlier in Section 3.2, we discussed shorter pauses that prevent speaker change, but longer pauses can also prevent a change if they happen at nongrammatical boundaries. We should recall Chafe's (1979) suggestion that long pauses are used for complex cognitive work. In other words, English speakers may utilize longer pauses for cognitive purposes rather than interactive purposes.

Pre-IU pause (sec)	Sp-change	No Sp-change	total
2-2.6	0 (0%)	5 (100%)	5
1	5 (41.6%)	7 (58.3%)	12
0.5	14 (27.4%)	37 (72.5%)	51
0.1	51 (20.4%)	198 (79.5%)	249

Table 8: Pause Length and Speaker Change in English (ECO#003)

Table 8 illustrates the relationship between pre-IU pause length and the likelihood of speaker change in the English data (ECO#003). Speaker changes occurred most frequently when the pause length was 1 second (41.6%) and less frequently after shorter pauses (27.4% after a 0.5-second pause and 20.4% after a 0.1-second pause). Interestingly, the longest pauses, exceeding 2 seconds, did not result in any speaker change. Instead, the same speaker maintained control of the conversational floor. Observe the following excerpt.

(15)

384	F3R	2	I=?	[0.7]
385	F3R	**2.6**	I don't know.	[0.4]
386	F3R	0.1	it's definitel=y like.	[1.1]
387	F3R	**2.6**	like cuz I= am (..) um,	[1.9]
388	F3R	0.5	I got stuck paying,	[1.3]
389	F3R	**1.1**	rent (..) um,	[0.7]

This speaker produced 6 IUs with three long pauses exceeding 1 second. These long pauses did not elicit a response because they occurred at points that were not syntactically complete junctures (Ford, Fox, & Thompson 1996). All five instances of pauses longer than 2 seconds in Table 8 follow this pattern. These extended pauses appear because the speaker is struggling to produce a coherent utterance.

In Section 4, I have presented the results of data analysis on IU length and pre-IU pause length in English and Japanese. The findings indicate that both IU length and pre-IU pause length are statistically significantly longer in Japanese. I argued that morphological differences account for the longer IU in Japanese, while the complexity of information that must be encoded in a clause explains the longer pre-IU pause length. Additionally, very long pauses trigger a speaker change in Japanese, but they prevent a change when they appear at nongrammatical boundaries in English.

5 Summary and Conclusion

In this paper, I examined the temporal aspects of IUs, including both the length of an IU and the duration of the pre-IU pause, using newly obtained semi-controlled natural conversation data from English and Japanese. The data analyses revealed that the median length of IUs is longer for Japanese (1.3 seconds) than for English (1 second). This difference, I argued, is due to the morphological differences between the two languages: Japanese words tend to be longer than English words. Regarding the pre-IU pause length, Japanese also shows a longer duration (median length of 0.3 seconds) than English (0.1 seconds). This is because, I argued, Japanese speakers need a longer pause duration to prepare to encode more pragmatic and discourse information within a single IU. Additionally, Japanese speakers use extended pauses for the purpose of speaker shifts. They may pause for a prolonged period to allow the addressee to take their turn. When the pause length approaches 1 second, more frequent speaker shifts occur. Conversely, English speakers use long pauses during the delivery of complex information to keep the floor.

Before concluding this paper, I offer several considerations for future research. First, pre-IU pause length and IU length may be culturally determined. While this may be the case, the cause of such cultural behavior still needs to be established. The cognition-focused and grammar-based explanation presented in this paper may offer key insights into this issue. Second, although I briefly touched upon genre differences, there is a clear distinction in the patterns of pauses and IU usage between narrative storytelling, such as the Pear Story, and interactive conversations. This paper focuses on the patterns observed in the latter genre. Further comparisons across different genres could provide deeper insights into the relationship between IUs and genre. Finally, since the Covid-19 corpus includes four additional languages besides English and Japanese, it is essential to examine other languages, such as Mandarin, Cantonese, Korean, and Thai, to elucidate both general and language-specific patterns of IU use across diverse languages.

The IU is a crucial unit of utterance in spoken language. Since their temporal behaviors are influenced by the speaker's cognitive processes and their interactional moves, a careful study of IUs can lead to a better understanding of how we think and communicate. I hope this study has contributed to advancing our research in this direction.

References

Barth-Weingarten, D. 2016. *Intonation units revisted: Cesuras in talk-in-interaction*. Amsterdam/Philadelphia: John Benjamins.

Beckman, M. E., J. Hirschberg, and S. Shattuck-Hufnagel. 2005. The original ToBI system and the evolution of the ToBI framework. *Prosodic typology: The phonology of intonation and phrasing*, ed. S.-A. Jun, 9–54. Oxford: Oxford University Press.

Biber, D. and S. Conrad. 2019. *Register, genre, and style*. Cambridge: Cambridge University Press.

Chafe, W. 1979. The flow of thought and the flow of language. *Syntax and semantics 12: Discourse and syntax*, ed. T. Givón, 159–81. New York: Academic Press.

Chafe, W. 1980. *The Pear Stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: ABLEX Publishing Corporation.

Chafe, W. 1988. Linking intonation units in spoken English. *Clause combining in grammar and discourse*, eds. J. Haiman and S. Thompson, 1–28. Amsterdam/Philadelphia: John Benjamins.

Chafe, W. 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing.* Chicago: University of Chicago Press.

Chafe, W. 2018. *Thought-based linguistics: How languages turn thoughts into sounds.* Cambridge: Cambridge University Press.

Clark, H. H. 1996. *Using language*. Cambridge: Cambridge University Press.

Coupé, C., Y. M. Oh, D. Dediu, and F. Pellegrino. 2019. Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances* 5(9): eaaw2594.

Croft, W. 1995. Intonation units and grammatical structure. *Linguistics* 33(5): 839–82.

Croft, W. 2007. Intonation units and grammatical structure in Wardaman and in cross-linguistic perspective. *Australian Journal of Linguistics* 27(1): 1–39.

Cruttenden, A. 1997. *Intonation.* Cambridge: Cambridge University Press.

Crystal, D. 1979. Neglected grammatical factors in conversational English. *Studies in English Linguistics*. Longman, eds. S. Greenbaum, G. Leech, and J. Svartvik, 153–66. London/New York: Longman.

De Ruiter, J. P., H. Mitterer, and N. J. Enfield. 2006. Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language* 82(3): 515–35.

Du Bois, J. W. 1991. Transcription design principles for spoken discourse research. *Pragmatics* 1(1): 71–106.

Du Bois, J. W., S. Schuetze-Coburn, D. Paolino and S. Cummings. 1992. *Discourse transcription*. Santa Barbara, CA: Department of Linguistics, University of California, Santa Barbara.

Durie, M. 1994. A case study of pragmatic linking. *Text-Interdisciplinary Journal for the Study of Discourse* 14: 495–530.

Enfield, N. J. 2017. *How we talk: the inner workings of conversation*. New York: Basic Books.

Enfield, N. J. 2025. The enchronic envelope. *Psychological Review.* Advance online publication. https://doi.org/10.1037/rev0000553

Ford, C. E., B. A. Fox, and S. A. Thompson. 1996. Practices in the construction of turns: The "TCU" revisited. *Pragmatics* 6(3): 427–54.

Givón, T. 2008. *Syntax: an introduction*. Vol. 1. Amsterdam/Philadelphia: John Benjamins.

Goldman-Eisler, F. 1967. Sequential temporal patterns and cognitive processes in speech. *Language and Speech* 10(2): 122–32.

Halliday, M. A. K. 1967. Notes on transitivity and theme in English: Part 2. *Journal of Linguistics* 3(2): 199–244.

Heldner, M. and J. Edlund. 2010. Pauses, gaps and overlaps in conversations. *Journal of Phonetics* 38(4): 555–68.

Himmelmann, N. P., M. Sandler, J. Strunk, & V. Unterladstetter. 2018. On the universality of intonational phrases: A cross-linguistic interrater study. *Phonology* 35(2): 207–45.

Inbar, M., S. Genzer, A. Perry, E. Grossman, and A. N. Landau. 2023. Intonation Units in spontaneous speech evoke a neural response. *Journal of Neuroscience* 43(8): 189–200.

Inbar, M., E. Grossman, and A. N. Landau. 2020. Sequences of Intonation Units form a ~ 1 Hz rhythm. *Scientific Reports* 10: 15846.

Iwasaki, S. 1993. The structure of the intonation unit in Japanese. *Japanese/Korean Linguistics 3*, ed. Soonja Choi, 39–53. Stanford, CA: CSLI.

Iwasaki, S. and H. Tao. 1993. A comparative analysis of the intonation unit in English, Japanese and Mandarin. Paper presented at the *63rd Annual Meeting of the Linguistic Society of America*, Los Angeles, January 9, 1993.

Levinson, S. C. 2016. Turn-taking in human communication–origins and implications for language processing. *Trends in Cognitive Sciences* 20(1): 6–14.

Levinson, S. C. and F. Torreira. 2015. Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology* 6: 731.

Matsumoto, K. 2003. *Intonation Units in Japanese Conversation: Syntactic, informational and functional structures*. Amsterdam/Philadelphia: John Benjamins.

Nagahara, H. and S. Iwasaki. 1994. Tail pitch movement and the intermediate phrase in Japanese. Paper presented at the *68th Annual Meeting Language Society of America*, Boston, January 6–9, 1994.

Ono, T., E. Yoshida, and M. Banno. 1998. It takes two to dance: The interactional determinants of NP intonation units with a marked rising intonation (*hangimonkei*) in Japanese conversation. *Japanese/Korean Linguistics 7*, eds. N. Akatsuka, H. Hoji, S. Iwasaki, S.-O. Sohn, and S. Strauss, 95–103. Stanford, CA: CSLI.

Park, J. S.-Y. 2002. Cognitive and interactional motivations for the intonation unit. *Studies in Language* 26(3): 637–80.

Pawley, A. and F. H. Syder. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and Communication*, eds. Jack Richards R.W. Schmidt, 191–225. London: Routledge.

Pellegrino, F., C. Coupé, and E. Marsico. 2011. Across-language perspective on speech information rate. *Language* 87(3): 539–58.

Pöppel, E. 2009. Pre-semantically defined temporal windows for cognitive processing. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364(1525): 1887–96.

Riest, C., A. B. Jorschick, and J. P. de Ruiter. 2015. Anticipation in turn-taking: mechanisms and information sources. *Frontiers in Psychology* 6: 62–75.

Selkirk, E. 1984. *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press.

Stivers, T., N. J. Enfield, P. Brown, C. Englert, M. Hayashi, T. Heinemann, G. Hoymann, F. Rossano, J. P. de Ruiter, K.-E. Yoon, and S. C. Levinson. 2009. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences* 106(26): 10587–92.

Tanaka, H. 1999. *Turn-taking in Japanese conversation: A study in grammar and interaction.* Amsterdam/Philadelphia: John Benjamins.

Tao, H. 1996. *Units in Mandarin conversation: prosody, discourse and grammar*. Amsterdam/Philadelphia: John Benjamins.

Templeton, E. M., L. J. Chang, E. A. Reynolds, M. D. Cone LeBeaumont, and T. Wheatleya. 2022. Fast response times signal social connection in conversation. *Proceedings of the National Academy of Sciences* 119(4): e2116915119.

Appendix

(A) Main transcription symbols used in this paper.

(.) = falling pitch contour/final sounding pitch contour; (,) = continuing pitch contour; (?) = rising pitch contour; (^) = rise-fall tail pitch contour for Japanese; (:) = lengthened segment for Japanese;

(=) = lengthened segment for English; (-) = cut-off syllable or word; (H) = inhale; (@) = laughter token; (<@ xx @>) = articulation with laughter

(B) Abbreviations used in this paper

1SG = first person singular
1SPL = first person plural
ASP = aspect
COM = comitative
COND = conditional
COP = copula
DM = discourse marker
GEN = genitive
HES = hesitation
INJ = interjection
NEG = negative
PAST = past
POL = politeness
PP = pragmatic particle
Q = question marker
QT = quotation marker
SE = sentence extender
TE = *te* form
TOP = topic

(C) Statistical tests

1. Comparing Means with Welch Two Sample t-test

- For Pre-IU lengths, the p-value was p < 2.2e-16, which is <0.05, indicating a **significant difference** between English and Japanese.
- For IU lengths, the p-value was p < 2.2e-16, which is <0.05, indicating a **significant difference** between English and Japanese.

2. Comparing Medians with Bootstrapping

- For Pre-IU lengths, the bootstrapped 95% confidence interval for the median difference was [0.1, 0.2]. Since 0 is not included in this range, we reject the null hypothesis and can conclude that Japanese and English IU lengths **are statistically different**.
- For IU lengths, the bootstrapped 95% confidence interval for the median difference was [0.2, 0.3]. Since 0 is not included in this range, we reject the null hypothesis and can conclude that Japanese and English IU lengths **are statistically different**.

3. Comparing Modes by Inspection

- The modes of Pre-IU lengths for English and Japanese are both 0.1.
- The modes of IU lengths for English and Japanese are 0.9 and 1.0, respectively.

(D) A list of frequently used words in English and Japanese identified by AntConc Program: (available from *Laurence Anthony's Website*)

ECO#001

Word	Lex cat	Sylla-ble	Rank	Freq
weird	adj	1	99	6
normal	adj	2	85	7
very	adv	1	41	14
all	adv	1	55	10
more	adv	1	65	9
much	adv	1	66	9
some	adv	1	69	9

really	adv	2	26	23
only	adv	2	73	8
over	adv	2	93	6
actually	adv	4	77	7
i	n	1	2	196
you	n	1	5	105
my	n	1	19	35
we	n	1	46	13
what	n	1	47	13
stuff	n	1	53	11
time	n	1	54	11
your	n	1	63	10
she	n	1	68	9
pizza	n	1	74	8
things	n	1	76	8
he	n	1	80	7
me	n	1	83	7
their	n	1	86	7
who	n	1	87	7
life	n	1	91	6
these	n	1	95	6
thing	n	1	96	6
way	n	1	98	6
people	n	2	29	21
something	n	2	75	8
pandemic	n	3	40	14
family	n	3	90	6
know	v	1	14	64
have	v	1	28	21
think	v	1	45	13
can	v	1	48	12
feel	v	1	64	9
get	v	1	71	8
has	v	1	72	8
go	v	1	78	7
guess	v	1	79	7
am	v	1	88	6
are	v	1	89	6
take	v	1	94	6
want	v	1	97	6
were	v	1	100	6
going	v	2	49	12
having	v	2	57	10

JCO#001

word	Lex cat	Syllable	Rank	Freq
あんまり	adv	3	55	2
すごい	adv	3	67	2
たぶん	adv	2	32	4
ちょっと	adv	3	33	4
また	adv	2	45	3
めっちゃ	adv	3	79	2
やっぱり	adv	4	46	3
今	adv	2	47	3
全然	adv	2	88	2
意外と	adv	4	91	2
本当に	adv	3	93	2
確かに	adv	4	94	2
結構	adv	2	52	3
あたし	n	3	100	1
おじいちゃん	n	3	60	2
お客さん	n	4	61	2
ゼミ	n	2	37	4
テレワーク(みたいな)	n	4	83	2
バイト	n	3	84	2
マスク	n	3	85	2
去年(の)	n	2	48	3
感じ(で)	n	2	50	3
時間	n	2	92	2
私	n	3	51	8
親	n	2	96	2
思います	v	5	49	3
思って	v	4	90	2
聞きました	v	5	95	2
言って	v	2	97	2