

Ranking Competencies of Oral Output: A Unit of Analysis for Low-Proficient L2 Speakers

Nobue Tanaka-Ellis^{1*}

¹ International Education Center, Tokai University, Kanagawa, Japan

* Nobue Tanaka-Ellis, E-mail: ellis@tsc.u-tokai.ac.jp

Received: August 7, 2017

Accepted: August 16, 2017

Online Published: August 22, 2017

doi:10.22158/selt.v5n3p560

URL: <http://dx.doi.org/10.22158/selt.v5n3p560>

Abstract

The evaluation of spoken language requires a rigid structure to make the analysis reliable, however, it is challenging to set the right criteria that work for different types of speech samples. Spoken language may be evaluated from different perspectives, depending on what to look at (e.g., intelligibility, Complexity, Accuracy, and Fluency). The former tend to concern a holistic evaluation of the spoken data, whereas the latter involves numeric measurements. Although they both analyse oral data, they do not seem to coexist in the same research field. Also, the research tools in both fields do not cater for evaluating data produced by speakers of low proficiency. This is because once a speech is labelled “low”, there are no additional classifications for further analysis. This paper, therefore, attempts to create those further categories. In this paper, intelligibility and existing methods for analysing spoken data in Second Language Acquisition (SLA) are reviewed to see if there are some overarching themes in teaching, assessing, and analysing spoken languages. Some of the issues from intelligibility and SLA are, then, delineated for designing a unit of analysis. The paper, finally, proposes the hierarchical C-unit, which is designed to deal with oral data produced by low-proficient speakers.

Keywords

low proficiency, intelligibility, unit of analysis, EFL, spoken language

1. Introduction

Achieving a good command of English is highly valued in Japan although the average English proficiency of Japanese people is, in fact, among one of the lowest in the developed countries and in Asia. The Japanese government lowered the starting age of English education from the middle school to the fifth-grade primary school from 2008 as a solution (Education First, 2017; Japan Times, 2015; Ministry of Education, Culture, Sports, Science and Technology Japan, 2014). Despite the Japanese government and educators' efforts to reverse the situation, it has not been successful (Japan Times, 2016). There may be myriads of factors contributing to why this situation shows little improvement,

however, one of the reasons may lie in the current research in second and Foreign Language (L2/FL) teaching or pedagogy. For instance, the current research methods or feedback systems may not cater for analysing the spoken data provided by very low-proficient speakers of English. Once learners are categorised under “low level” for their proficiency, they may not receive detailed feedback on their L2 skills although there are numerous factors that learners need to work on in order to improve their English speaking skills. Spoken data or samples may be assessed or evaluated from a few different aspects, such as intelligibility and Complexity Accuracy and Fluency (CAF). The former looks at spoken language holistically, and the latter is a tool for analysing speech data to determine complexity, accuracy, and fluency for numerical measurements. In the next section of the paper, I explore how intelligibility may relate to L2 and FL evaluation and how current measurements can serve low-proficient speakers.

2. Intelligibility

The notion of the term “intelligibility” has been changing over years, and according to Nelson (2008), its paradigm has shifted dramatically after the publication of Smith and Rafiqzad’s paper in 1979. In Nelson’s paper published in 2008, he traced changes in the notion of intelligibility in a chronological manner stating that intelligibility initially perceived as phonetic and phonological criteria. The studies published before 1979 mainly focused on phonetic and phonological differences between the English spoken by non-native speakers and native speakers from the regions categorised as the Inner Circle (Kachru, 1985, 1986), such as the US, the UK, Canada, Australia, and New Zealand. For instance, the greater the difference in the phonological properties between a non-native speaker’s utterances and those produced by the speakers from the Inner Circle, the non-native speaker was regarded more unintelligible. In Smith and Rafiqzad’s research (Nelson, 2008), intelligibility was studied as accessibility of pronunciation which results changed the perception of intelligibility from native speakers’ standpoint to comparisons of different variety of Englishes. Nelson describes this paper “ground-breaking” (2008, p. 301) as it revealed that among different speakers of English, the native speaker’s speech was always rated as one of the least intelligible samples and the Japanese sample was always within the top five to over 1,300 subjects across eleven countries in Asia and the Pacific region. Nelson then mentions that a speaker can be intelligible but uninterpretable at the same time due to the differences in deep cultures between the two native speakers that share the same language (e.g., a US speaker vs. an Australian speaker). In the same paper, Nelson (2008) also introduces a few different definitions of intelligibility mentioned by different researchers, such as intelligibility as comprehensibility of syntax and semantics, comprehensibility, accentedness, interpretability, or combinations of these factors.

Intelligibility is a complex issue and it has been argued from different perspectives. For instance, Berns (2008) claims that intelligibility in English as a Lingua Franca (ELF) and World Englishes (WE) are slightly different. While both ELF and WE agree that non-native speakers of English are right to their

own pronunciation and the top-down imposition of the correct accent(s) by native speakers should be dismissed, they disagree on the position regarding intelligibility of Expanding Circle speakers. Unlike WE, ELF marginalises NSs in international communication in two different points: 1) ELF claims that each non-native speaker is responsible for gaining the confidence in their accents and not let them be the obstacle for producing intelligible speech and, 2) the target audience or interlocutors are international English speakers (i.e., any combinations of speakers: native or non-native). Drawing from the aforementioned arguments and the results of Smith and Rafiqzad's study, if Japanese speakers are able to improve their English skills, they may have an advantage in terms of improving their intelligibility. Unfortunately, intelligibility is a very elusive construct and its methods of measurement and evaluation have not seemed to be quite established.

Assessment of intelligibility is traditionally related to speech pathology or first language development and disorder (Levy, Leone, Moya-Gale, Hsu, Chen, & Ramig, 2016; Pennington, 1999; Perez, 1983; Pickett & Pollack, 1963), therefore, the research foci are different from that of in the field of second or foreign language education. Much of the research in intelligibility in second and foreign language research is related to interpretability, comprehensibility, and pronunciation (e.g., Anderson-Hsieh & Koehler, 1988; Anderson-Hsieh, Johnson, & Koehler, 1992). Anderson-Hsieh, Johnson and Koehler's (1992) study in particular, focuses on how non-native speakers' pronunciation deviated from the Englishes used in the Inner Circle. The raters in this study were all English as a Second Language (ESL) teachers in the US. The papers like this clearly adopted the WE view of intelligibility whose research evolves around the view from the Inner Circle speakers' perspectives, which may be problematic when this is applied to language teaching. In Sewell's paper (2010), different types of intelligibility studies are introduced, which are mainly focused on phonological analyses of both native and non-native speakers. The paper discusses the divide that exists in approaches to intelligibility studies that were conducted from an ELF perspective, or a WE perspective. This divide can potentially be problematic in formulating methods for measuring intelligibility universally and connect the outcome of the research back to English education and pedagogy. In fact, Rajadurai (2007) claims that intelligibility studies and the outcomes lack consistency, which may be due to their diverse definitions, methodologies, and samples used in their investigations.

Coetzee-van Rooy (2009) conducted a study on how English speakers in the Outer Circle (e.g., regions where English is used as second/official language: India, Singapore, and the Philippines) perceived speech samples collected from three people from the Expanding Circle. The speech samples through structured interviews with three South Koreans were evaluated by 18 university students in South Africa on: perception and attitude towards the speakers, intelligibility, comprehensibility, interpretability, and ease in understanding. In another example of speech perception, Tan (2015) investigated how stress in bi-syllabic and tri-syllabic words spoken by Singapore English speakers were perceived by Singapore English speakers and different native speakers of Englishes (American, Australian, and British). The study revealed that native English speakers in the Inner Circle perceived

the locations of syllables differently from that of the Singapore English speaking participants. Also, the Singaporean speakers showed sensitivity to their own English variety in differentiating stress, which was not seen in speakers from other regions. Tan maintains that Singapore English speakers should be regarded as native speakers of their own English variety, rather than being categorised as non-native speakers of English used in the Inner Circle. Interestingly, in another study of perception by Rivers (2011), different results were reported.

In Rivers's study, 48 Japanese university students' attitudinal responses towards 10 speakers with different accents in English (1 Japanese speaker, 7 non-native speakers from different Asian countries, and 2 native speakers from the US and the UK) were looked at. He found that more than half of the Japanese participants thought that the Japanese speech sample was spoken by an American speaker, the American sample was spoken by a British speaker, and the Chinese sample was provided by a Japanese speaker. Unlike Tan's Singaporean study, the Japanese participants did not show any sensitivity towards the Japanese accented English. In this paper, Rivers revealed that the Japanese speakers valued the varieties of English represented by the native English speakers high; on the other hand, the varieties of English spoken with Asian accents were perceived negatively. From his study, Rivers suggests implications of English language teaching in Japan and that Japanese English learners are trapped in their inferiority complex. This complex is driven by a failure to achieve the unrealistic standard of native-like English imposed by their government, which prevents them to be open to the true diversity of English language. The negative notion of Asian accented of English varieties in Japanese learners needs to change before they can achieve a realistic goal that English is an international language. These types of research can certainly provide a great insight into how intelligible speakers of the Expanding Circle are to other non-native speakers or to native speakers. Only recently, studies like Tan's (2015) and River's (2011) started to examine the standpoint of non-native speakers of English, however, they have not been able to suggest any methods how non-native speakers should study to be more confident in their English as a Lingua Franca. These studies represent the very situation that Berns (2008) suggests in terms of ELF usage that non-native English speakers need to be confident in their own accents and not let them be the obstacle in communication. As traditional studies in intelligibility inherently focus on the differences between native and non-native accents and are still prevalent, as in Anderson-Hsieh, Johnson and Koehler's (1992) study, it seems that the ELF-WE divide would not dissolve anytime soon. Therefore, it would take a while to set the standard for determining what is intelligible and what is not from the ELF perspective. Another problem with incorporating the view of ELF into the classroom is that there is no universal definition of how intelligibility can be measured (Pickering, 2006). Nevertheless, since intelligibility is regarded as a prerequisite for successful communication (Sewell, 2010), constructing the means to evaluate and provide training for English learners to improve their intelligibility seem to be a logical step to take in the field of English as a Second or Foreign Language (ESL/EFL) teaching.

2.1 Intelligibility and Instruction

According to Jenkins (2004), more and more teaching materials have non-native speakers to expose learners to a variety of accents to emphasise that English is an international language. However, the English spoken by non-native speakers in these materials are intentionally made intelligible. Jenkins states that this is more for the teachers to raise awareness about English as an international language, rather than teaching students the accented pronunciation. Despite this recent change in English instruction, many teachers and learners prefer to target for native-like form of English although they also believe that learners should achieve international intelligibility and accented English is acceptable. Munro and Derwing (1995), in fact, found that accented English spoken by Cantonese, Japanese, Mandarin, Polish and Spanish speakers, did not influence the comprehensibility of speech. Similarly, in their later study (Munro & Derwing, 1999), 10 proficient Mandarin speakers and two native speakers of English were evaluated for intelligibility and comprehensibility. They again found that foreign accent did not interfere with comprehensibility. They further found that speakers who made grammatical mistakes tended to have pronunciation errors. Therefore, grammatical errors may lead to reduction of intelligibility, not due to foreign accents. Orikasa (2016) conducted a study on how Japanese speakers perceived intelligibility of speakers from different backgrounds (one female and one male speaker each from the US, China, Korea, and Vietnam). The researcher found that speech rate was the main contributor of unintelligibility. The speakers with the lowest intelligibility were the female US speaker and the male Vietnamese speaker. Orikasa concludes that the language pedagogy needs to be inline with the current diversity of English users. Finally, Thomson and Derwing (2015) reviewed the papers on pronunciation training given in the classroom and computer-assisted pronunciation teaching modes. While pronunciation training can be highly effective in the classroom setting, the time constraint was the biggest problem. Computer-assisted mode, on the other hand, has an advantage in this respect that learners can access to pronunciation training as often as they want. According to Thomson and Derwing, the studies on pronunciation training lacked the ecological validity as each study only focused on limited aspects of pronunciation (e.g., improvements on one or two phonemes), therefore, the results need to be interpreted with caution. Drawing from the papers on intelligibility, even the stakeholders in English education have mixed opinions and, therefore, the conflicting views seem to be manifested in the direction of English teaching and research that they are still contemplating on the target outcome of English education.

2.2 Assessment

Smith and Rafiqzad (Berns, 2008, p. 333) suggest that “a true evaluation of speakers’ comprehensibility should be judged by both native and non-native speakers” and this applies to native speakers as well. Regarding assessment related to English as a Lingua Franca and World Englishes, Elder and Davies (2006) proposed two models for ELF. According to them, ELF is understood in four ways: 1) English is used where some of the participants are non-native speakers, 2) English is used where all the participants are non-native speakers and from different language backgrounds, 3) English is used where

all the participants are non-native speakers and speak the same or similar first language, and 4) new code (new form of English) based on standard English is used among non-native speakers. Based on the first three definitions, they proposed that an ELF assessment should look like that of the standard English test, such as SAT but cultural specific terms and context should be taken out from the test. Also, interlocutors in these tests should be either examiners or other test-takers who are proficient non-native or ELF speakers, so that they understand what the candidates are going through and able to adjust the communication style. Raters, both native and non-native speakers, need to be trained to ensure that they do not penalise culturally specific use of terms, and involve ELF users to check the borderline is low enough for proficient ELF speakers are able to pass the test. In the second model, Elder and Davies suggest that in case of assessing speakers who use a new code, for instance Hong Kong or Singapore English, the new form of English should be used, instead of standard English because they may have limited exposure to standard English. Davies (2009), however, claims that the acceptability of WE, including ELF, seems to be driven by political factors rather than applied linguistic realities and if WE varieties are accepted in assessments, then non-standard WE varieties would be excluded. He goes on and states that deciding which variety of English is to be tested is a language-planning decision that contributes to standardisation of WE, however, the decision can be made only when the speech community agrees on the criteria what is the direction of the WE. These papers exhibit that assessing WE or ELF involves myriads of factors that are not limited to linguistic elements, and all these factors need to be ironed out before implementing a test that accommodates WE and ELF users. This would be one of the main reasons why a set of universal principles and pedagogies for teaching and assessing intelligibility are yet to exist.

3. Analysing Spoken Data in Applied Linguistics

In the field of applied linguistics and SLA, oral output has been analysed from numerous aspects. As in intelligibility research mentioned in the previous section, pronunciation is well researched. Fluency is another aspect of speech that has been widely looked into (e.g., Lennon, 1990; Derwing, Munro, Thomson, & Rossiter, 2009; Bosker, Pinget, Quen é Sanders, & de Jong, 2012). In studies of fluency, there are two main methods of investigation. In Derwing, Munro, Thomson and Rossiter's (2009) study, native Canadian English speakers were used as raters to rate Mandarin and Slavic language speakers' English by using a 9-point scale (1: extremely fluent to 9: extremely dysfluent). Although all the raters had linguistics background and achieved high interrater reliability, this type of rating method relies on raters' perception and it is beneficial for grasping the holistic view of each speaker's fluency. On the other hand, Bosker, Pinget, Quen é Sanders and de Jong (2012) took a different approach in analysing fluency. They analysed fluency by measuring and calculating speed (mean length of syllables), breakdown of fluency (number of silent pauses, filled pauses, and mean length of silent pauses), and repair (number of repetitions and number of corrections) so that numerical measurements can be obtained.

Lennon (1990) suggests two senses of fluency in an EFL context: a broad sense and a narrow sense. The former refers to the global language performance and often is the highest point of a language proficiency scale, and the latter points at a component of oral proficiency among other factors, such as grammar and vocabulary knowledge. Apart from analyses of pronunciation and fluency, one of the most prominent methods of analysing language data is Complexity, Accuracy, and Fluency (CAF) that looks at grammatical complexity, grammatical accuracy and fluency in speech, which has been employed by many researchers (e.g., Yuan & Ellis, 2003; Iwashita, Brown, McNamara, & O'Hagan, 2008; Skehan & Foster, 2008). CAF may be employed in many language studies, however, it is not without issue. Norris and Ortega (2009) summarise how complexity was measured in different studies. In their paper, they tabulated units of analysis used in different studies (1979 to 2007) that employed CAF. They found that each study used different methods of measuring complexity, that is, different units of analysis were found and some of them were not used in SLA. They then focused on 16 more recent studies from 2001 to 2008 to find out how complexity was measured. Compared to earlier studies, newer studies employed relatively similar methods from each other, but even so there were three main units (T-unit, C-unit, and AS-unit) of analysis found in these papers.

3.1 Unit of Analysis for Spoken Data

In Foster, Tonkyn and Wigglesworth's (2000) study, units for measuring spoken language were investigated. They found that more than half of the articles (44 out of 87) that they reviewed in four major academic journals in Applied Linguistics and SLA fields provided no definition of unit of measurements. The authors also found that the definitions of the units provided in the other half of the articles varied greatly in how much detail provided in defining these units. Three loosely categorised units are presented in their paper: mainly semantic units, mainly intonational units, and mainly syntactic units. Under "mainly semantic units", for instance, three different studies are categorised but they all have different names and definitions from each other. The definitions in this category rarely stand alone and usually accompanied by grammatical and intonational criteria, due the difficulties in relying only on semantic properties. With the "mainly intonational units" category, the authors mention that the studies listed under this category used problematic methods especially using intonation and pauses may not function as a unit when measuring the samples produced by L2/FL or dysfluent speakers that can have varieties of speech patterns. In the "mainly syntactic units", a unit is defined as sentence, idea unit, T-unit, or C-unit. The first definition, sentence, is problematic especially in spoken language, the authors do not elaborate on this matter. The idea unit is defined as a clause constructed by pre- and post-V elements. The most well known unit, T-unit, is used for both spoken and written data, which is generally defined as an independent clause or a dependent clause are counted as T-unit. This is, however, unworkable when data of dysfluent speakers are analysed because it would contain many incomplete clauses. C-unit seems to be the solution to the problem found in using T-unit. For example, Iwashita (2001) explains that C-units provide communication value through words, phrases, and sentences, which can be both grammatical and ungrammatical. Although there are several definitions of

C-unit, grammatical predications or a meaningful word that serves as an answer to a question, like “yes” or “no” are regarded as C-units. Rimmer (2010) states that the core of the problem is deciding what is grammatical or not in measuring grammatical complexity. For instance, he gives an example by showing a phrase like “How awful he was!” which can be expressed as “How awful!” (p. 501) in an interaction, however, a problem arises when this phrase is being analysed for grammatical correctness that if this shorter utterance can be treated as grammatical or ungrammatical. If T-unit is used for analysing this phrase, it would not be regarded grammatical, on the other hand, C-unit would allow it to be counted as one unit. Finally, Foster, Tonkyn and Wigglesworth (2000) proposed the AS-unit, a syntactic unit, for segmenting oral discourse which is defined as a single speaker’s unit consists of an independent clause or sub-clausal unit, with subordinate clauses. They claim that unlike T-unit, AS-unit allows C-unit like utterances that may not include a verb.

These units are, however, often unsuitable for dealing with data collected from dysfluent or low-proficient L2 users because once the data is under the threshold of “low” or “extremely low”, no further analysis is conducted. This paper concerns this type of data that falls under such categories because a new unit of analysis capable of analysing the speech samples of low-proficient speakers is needed. It is, therefore, important to further analyse C-unit because low-proficient speakers often produce series of C-units in order to put their meaning across to their interlocutors. This would help understand the characteristics of many low-proficient and dysfluent speakers who exist in language classrooms in the Expanding Circle (i.e., Japan).

In an attempt to create a new unit of analysis, it would make sense to revisit some of the issues that were discussed in the previous section; especially, the issues related to intelligibility. The following is the list of the main issues about intelligibility that were discussed earlier.

- Foreign accents are not the major contributors to unintelligible speech.
- Speech rate (fast) can cause unintelligibility.
- Grammatical error may lead to pronunciation errors.
- English as a second or foreign language speakers need to be confident in their pronunciation.
- Native speakers are not often the most intelligible speakers to non-native speakers, and non-native speakers may be more intelligible to other non-native speakers.
- Teachers and learners both prefer native-like oral output.
- Pronunciation training is effective but current studies lack ecological validity.
- Both teachers and learners should acknowledge a diverse variety in English.

From the list above, all the components in CAF may not need to be incorporated for very limited users of English. Complexity may be analysed by looking at C-units, rather than T-units or AS-units. Accuracy may be determined by looking at grammatical features in speech samples and coherence in an utterance. As C-unit is already defined as a smallest meaningful word or phrase, determining if the C-unit is grammatical or ungrammatical could be a good way to grasp the levels of low-proficient speakers. Fluency may be determined by number of words uttered in a certain time frame. However,

low-proficient speakers may have problems with uttering a sentence or a C-unit with a fast enough speech rate that can be captured as one chunk of unit because they often have long pauses between words or even within a word. Pronunciation may be looked at, however, it is challenging to assess pronunciation of already dysfluent speakers and what is acceptable and what is not at this level may primarily be raters' preference, and therefore, this component may not be as important as accuracy of syntax. Based on these assumptions, I propose the hierarchical C-unit for analysing spoken data collected from speakers with English with low proficiency.

4. Hierarchical C-unit

The current units for analysing speech data are not able to analyse the speech samples that are uttered by very low-proficient speakers. The unit I have adopted to create the hierarchical C-unit (HC-unit) is based on the C-unit (see Iwashita, 2001), which can also be defined as an independent clause with its modifiers (e.g., subordinate clauses) that cannot be further divided into smaller units without obscuring the basic message. Because a C-unit is usually considered as the smallest meaningful word, a phrase, or a sentence that can be both grammatical and ungrammatical, erroneous utterance can be analysed if it makes sense in the context where the utterance that contain C-units are spoken. However, it is important to analyse C-units beyond counting the number of units in a certain length of an utterance. For the teachers and researchers who are usually working with L2 or FL speakers with very limited oral skills, it is beneficial for them to know the type and quality of learners' C-units. The HC-unit was designed to look at spoken data collected from English speakers from the Expanding Circle because this is where a large number of speakers with low English proficiency exist, therefore, the speech samples of speakers of English as a foreign language are the data I refer to in this paper, rather than that of from the speakers of English as a second language.

4.1 The Structure of the Hierarchical C-unit

This section explains the structure of the HC-unit and demonstrates how it can be used to analyse actual data. The following figure (Figure 1) exhibits the structure of the HC-unit.

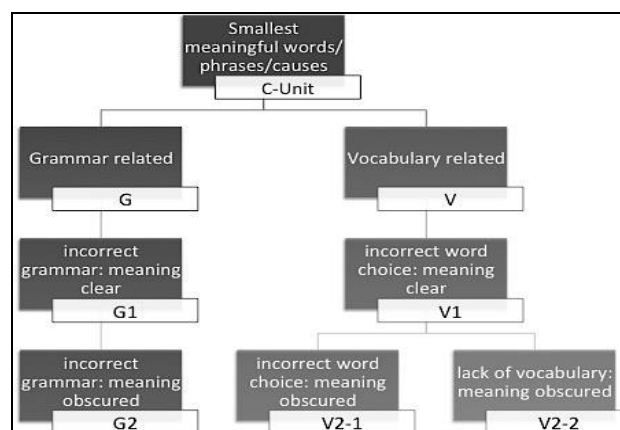


Figure 1. The Hierarchical C-unit

The top tier indicates a C-unit, which can be broken down into two main categories: Grammar related (G), and Vocabulary related (V). If a C-unit does not contain any errors, the categorisation would end here. However, if the C-unit contains errors, then the errors can be categorised in two different levels. If the error is not serious and it does not confuse the message to be put across, it can be put under G1 or V1. If the error is more serious and it confuses the message to be conveyed, then it can be sorted under G2 or V2-1/V2-2. It is common for language learners to choose incorrect words or expressions, or exhibits difficulties in expressing themselves, therefore, V2-1 and V2-2 can differentiate these instances. V2-1 may be used when a speaker chooses a wrong word or phrases for the context, however, V2-2 may include the situation when a speaker replaces a L2 word to his/her native language or a partial word/phrase. Unlike other methods of analysing oral data (e.g., CAF), analysing complexity by calculating mean length of utterance, T-unit, or C-unit, etc. is extremely challenging. The hierarchical C-unit's categories (G or V) and correctness (levels 1 or 2) can provide some basic information of accuracy and complexity.

The spoken data used in these samples are collected from Japanese university students with science and engineering background, who attended a first-year compulsory English course. The speakers were all regarded as low-proficient speakers of English as their TOEIC scores ranged from 180 to 300, or CEFR A1 to A2 equivalent. There are a few ways of collecting samples from FL speakers. For example, recordings can be obtained through a few different ways; interviews, monologues, dialogues, or tasks. The following samples were collected through a task as a part of course assignment. The learners listened to interviews about how people in the UK feel about Christmas on BBC Learning English (2013). BBC provides downloadable audio files and transcripts for ESL/EFL learners. The audio file was uploaded to the university's Learning Management System (LMS), together with the task instruction. In the task, the learners were to record their opinions about the interviews of three UK speakers. The LMS allowed the person with an administrative role to set the maximum number of time each student could listen to the uploaded materials. In the task instruction, they were asked to record their own opinions about the three British speakers and upload MP3 files to the LMS. This is an easy way to collect speech samples, however, it is hard to control the length of the speech because some speakers may only manage to speak one sentence or phrase and some of them may be able to utter a few sentences. There is always a risk of students reading their own scripts, however, it is highly unlikely that TOEIC 180-300 speakers can suddenly speak like higher-level speakers. The topic of interview or tasks can also influence the length of speech greatly and the interlocutor's FL proficiency may deeply affect the outcome of samples, too. This is because the speakers with low proficiency usually have very limited knowledge of vocabulary and depending on their history of learning the target language, the range of vocabulary they can recognise and use may vary. Similarly, because these learners at this level are often unaware of the mistakes that their interlocutor makes and they often copy each others' expressions, if two low-proficient speakers engage in a dialogue, therefore, as mentioned in Elder and Davies's (2006) paper, it is important to choose an experienced interlocutor for assessment,

which applies to data collection as well.

In the following examples, a one-second pause is represented by (.), and different numbers in brackets indicate actual length of pauses in seconds. A vertical slash | indicates a boundary of C-unit. Each new C-unit is shown in a new line and the parts that are not counted as a C-unit are attached to the end of the previous C-unit. The hierarchical categories are in < >. Correct (or acceptable) C-units are not annotated.

1. Utterance with no serious mistakes: G1 and V1

a) 1 C-unit, 3 mistakes

|I agree with woman <G1> (.) because her opinion is interesting, very compact <V1> and easy to (.) image <G1> Christmas|

Example a) contains one C-unit. The first half of this C-unit can be analysed as G1 because the missing article “the” does not affect the meaning. In the subordinate clause contains V1 [very compact] and G1 [image].

b) 1 C-unit, 3 mistakes

|I agree with the second man because as for <G1> the time to spend with a family <G1>, it is important for anyone <V1>|

This sample was a little more fluent than other samples, however, it was short. The expression “because as for...” is awkward and it may be labelled as <G1>, however, the error is not serious and meaning is still clear. The V1 label indicates that “anyone” is a wrong word choice, which again, is not serious and the meaning of this utterance is not obscured. The next two examples show more serious errors in the utterances. Example c) has five C-units, however, the vocabulary level is not as good as the examples in the previous examples.

2. Utterance with a serious grammatical mistake: G2

c) 5 C-units, 4 mistakes

|I think that I agree with their opinion <G1> because I (2) *alway* [always] happy <G2>| (6) I (2)

|I think that I like Christmas| (2)

|Christmas is events <G1>|

|I like events| (2)

|Events is very (3) good <G1>|

The first C-unit contains an example of G2. The speaker probably wanted to say that he or she is always happy when the Christmas day comes, however, the coder or evaluator needs to make a deduction of the meaning from the context.

3. Utterance with serious vocabulary mistakes: V2-1

d) 4 C-units, 5 mistakes

|I agree with a woman <G1> because three of us <V2-1> are *aplipliate* [appropriate] <V2-1> for Christmas| (.)

|I *shink* [think] Christmas is a good event| (.)

[It is *posshibulu* [possible] to spent <G1> only family <G1>|

[It is *posshibulu* [possible] to *depun* [deepen] the relationship if there is a *laber* [lover]]

The example d) has two long sentences. However, in the first sentence, the speakers says I agree with “a woman” then he goes on to elaborate the reason by saying “three of us” <V2-1>. The coder can only guess if the speaker wanted to say “three of them” (the British interviewees) or “those three opinions”, or something different. Also, “are appropriate” makes it more complicated because it is unclear what this is referring to.

In the following section, analyses of different samples are demonstrated. This data was obtained from the similar type of students who were Science and Engineering undergraduate students. The task was not a part of their assessment and students were asked to provide their speech samples as research participants of a different study. This was a timed speech and students were asked to describe pictures as much as they could. The pictures were on a PowerPoint file and each picture was shown to them for 30 seconds. After 30 seconds, there was a five second gap with an empty slide, and then a different picture automatically appeared. There were 12 speakers and 8 pictures in total.

4. The same stimulant different outcomes

e) 4 C-units, 2 mistakes

|She is a woman|

|She is a beauty <G1>|

|She is lie <G1> on the sofa|

|Her hair is black| (3) She is (.) she is

f) 3 C-units, 3 mistakes

|She’s wearing a red clothes <G1>|

|She lie <G1> on a black sofa|

|She black hair <G1>| and ah she is an

g) 2 C-units, 1 mistake

|One woman is on a black sofa| Hmm

|Her is red dress <G1>| Yes. One woman is

The above examples show that three different students were describing about the same picture. The amount of information they put into their 30-second utterances is different. In e), “she is a beauty” may not be a mistake, however, considering the level of their target language, it is unimaginable that the sentence was uttered in the same way that more advanced speakers would. All the C-units with mistakes were not serious, however, this method of data collection can limit their speakers to have short sentences.

5. Higher proficiency

h) 4 C-units, 5 mistakes

|She <V1> is soccer <G1>| She ball

|He is shooting a ball|

[Ground <G1> is green|

[The keeper <V1> is (.) defend <G1>|

i) 5 C-units, 4 mistakes

[The man who is laughing is sitting a sofa <G1>|

[Maybe she <V1> is only ah lonely|

[He watching <G1> TV|

[Comedy|

[This room has three book shel ah book shore <V1>|

j) 5 C-units, 2 mistakes

[Ah she, this is woman <G1>| She is ah to be eee

[She is a soccer team owner|

[They are winning a tournament cup| Ah

[They are ah so very <G1> happy|

[And congratulations!| They are ah...

These three examples, h) to j) show some typical utterances recorded from the speakers with higher proficiency than other students. Note that these students were still in the basic-level class whose TOEIC scores were under 300. In h) and i) the speaker mixed up the gender usage, however, because the task is picture description, it helped the coder to understand whom this speaker was referring to. If the speaker made the same mistake in the other task, it would have been confusing and the meaning can be obscured. In i), the fourth C-unit, “Comedy”, is a valid example of a fragmented phrase which can be regarded as a C-unit. Similarly to i), the last C-unit from j) is fragmented, however, sometimes when speakers get excited while speaking, they talk to the pictures like in this example.

6. Lower proficiency

k) 4 C-units, 5 mistakes

[The man (.) play <G1> soccer|

[The man kick <G1> goal|

[The man shirt wear red <G1>|

[The man is handsome|

[The man is good job soccer <G1, V1>|

l) 2 C-units, 4 mistakes

[This animal is (.) wol, (2) wolf <G1>|

[One, two, three, four, wolves is <G1> (2) song. Song. <G1, V1>|

m) 2 C-units, 0 mistake

[A man (2.5) is (.) on a ship| (.) And uu (3) Heee (2)

[He fished a very big (1.5) tuna| (.) and (3) and (.) back

The above examples show speakers with lower proficiency, compared to other speakers in the data set. The sentences are simpler than the examples in 5, and they used fewer and simpler words. Note that

these utterances produced within 30 seconds, therefore, they also spoke each word slower than the previous examples and had longer pauses and more self-corrections, which resulted in a fewer C-units. In this data set, because the time for each descriptive task was set, the number of C-units within the time limit is a good indication of how fluent the speakers were. However, it is difficult to rank speakers with very different speaking styles like in k) and m). In k), for example, the speaker uttered 5 C-units with multiple errors while m) produced only 2 C-units with many long pauses. In terms of complexity, k) only used present tense with many mistakes, whereas m) may only uttered two C-units, they were without obvious mistakes and in the past tense, which is considered slightly more advanced than using the present simple. When comparing utterances like these, researchers need to be clear about what they are looking at. If fluency or amount of information per utterance is more important, k) may be a better speaker, however, if syntax is more important as discussed earlier in the intelligibility section, then m) may score higher.

In this paper, I have introduced the hierarchical C-unit with two different types of data, both collected from learners with TOEIC 180 to 300. As of 2016, the mean TOEIC score in Japan was 516, which ranked 41st out of 49 countries worldwide (ETS, 2016). The country with the lowest mean score was Indonesia with the score of 397. Therefore, the data dealt in this paper was even lower than the lowest mean score that verifies the purpose of using the data I used for presenting the HC-unit. The two different sample data sets showed that methods in collecting data brought out totally different types of utterances although they were both collected from EFL learners from the faculty of the same university and very similar TOEIC scores.

There are a few issues to consider when evaluating and analysing spoken data of low-proficient speakers and intelligibility. For instance, speaking with correct syntax may be important because foreign accents do not affect intelligibility. However, what is the difference between foreign accents and mispronunciation? The example d) was presented with speaker's very heavy accent. The words "think" and "possible" may be acceptable to be pronounced as "shink" and "posshibulu", however, how about "apliliate" and "labeer"? As in Tan's (2015) paper, Singapore English speakers are sensitive towards their own pronunciation, however, in Orikasa's (2016) paper, the Japanese students thought that the speech sample spoken by a Japanese speaker as American English. This indicates that Singapore accent is established and Singapore English speakers speak in more or less the same way. English speakers in Japan, on the other hand, are probably not used to using English in every day situations, and therefore, they do not have an identity of speakers of Japanese English. EFL speakers in Japan may have a certain tendency in their English pronunciation, however, unlike Singapore English speakers, they do not possess the same norm and identity in their English, therefore, their pronunciation may differ greatly between speakers. The finding in Munro and Derwing's (1999) study suggested that speakers who made more syntactic errors, they tended to mispronounce words more frequently. I believe this is true because the accent in Singapore English is not due to syntactic errors and they know exactly what they mean. However, considering the fact that the speaker is a basic-level EFL user who

made multiple syntactic errors, “*apliplate*” is more likely a mispronunciation than an accent because the language identity may only be achieved through sharing the same accent and the its users feel that it is their language. In this sense, the acquiring grammatical knowledge is more important than fixing their pronunciation because the learners need to know exactly what they mean.

5. Conclusion

This paper attempted to incorporate some main ideas and arguments from two research areas that are closely related but not in a close collaboration into designing a new unit for analysing spoken data produced by very limited users of English as a foreign language. It is clear that the definition given for the hierarchical C-unit may not be applicable to different types of data and researchers may need to make some adjustments for the data type they have in hand. Because C-units can take so many different forms, counting the number of C-units alone is not enough for analysing the lower level learners because the quality and the complexity of a C-unit is what researchers and educators should look into more closely. Transcribing and finding C-units in a large quantity of spoken data are already time consuming, however, systematically labelling types of C-unit, searching for errors, and ranking the severity of errors may provide a good starting point to understand the kind of data that researchers are dealing with, and decide how they want to analyse from this point. In this respect, those two data sets have shown how differences in data format can influence in analysing them and how researchers need to adjust the way they analyse or even the method of data collection to ensure that they obtain the information they wish to work on. Either the field of intelligibility in World Englishes and English as a Lingua Franca or Second Language Acquisition, whatever the method is helping language learners with perfecting their grammatical skills seems to be the pivotal factor for improving their overall intelligibility.

References

- Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker. *Language Learning*, 38(3), 561-613. <https://doi.org/10.1111/j.1467-1770.1988.tb00167.x>
- Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of non-native pronunciation and deviance in segmentals, prosody, and syllabus structure. *Language Learning*, 42(4), 529-555. <https://doi.org/10.1111/j.1467-1770.1992.tb01043.x>
- BBC Learning English. (2013, December 19). 6 Minute English: Festive feelings. *BBC Learning English*. Retrieved December 20, 2013, from http://www.bbc.co.uk/worldservice/learningenglish/general/sixminute/2013/12/131219_6min_festive_feelings.shtml
- Berns, M. (2008). World Englishes, English as lingua franca, and intelligibility. *World Englishes*, 27(3/4), 327-334. <https://doi.org/10.1111/j.1467-971X.2008.00571.x>

- Bosker, H. R., Pinget, A.-F., Quen é H., Sanders, T., & de Jong, N. H. (2012). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, 30(2), 159-175. <https://doi.org/10.1177/0265532212455394>
- Coetzee-van Rooy, S. (2009). Intelligibility and perceptions of English proficiency. *World Englishes*, 28(1), 15-34. <https://doi.org/10.1111/j.1467-971X.2008.01567.x>
- Davies, A. (2009). Assessing world Englishes. *Annual Review of Applied Linguistics*, 29, 80-89. <https://doi.org/10.1017/S0267190509090072>
- de Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, 61(2), 533-568. <https://doi.org/10.1111/j.1467-9922.2010.00620.x>
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition*, 19, 1-16. <https://doi.org/10.1017/S0272263197001010>
- Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 development. *Studies in Second Language Acquisition*, 31(4), 533-557. <https://doi.org/10.1017/S0272263109990015>
- Education First. (2017). *The world's largest ranking of countries by English skills* (6th ed.). EF English Proficiency Index. Retrieved July 15, 2017, from <http://www.ef.edu/epi/>
- Elder, C., & Davies, A. (2006). Assessing English as a lingua franca. *Annual Review of Applied Linguistics*, 26, 282-301. <https://doi.org/10.1017/S0267190506000146>
- ETS. (2016). *Report on Test Takers Worldwide: The TOEIC Listening & Reading Test*. Retrieved August 6, 2017, from https://www.ets.org/s/toeic/pdf/ww_data_report_unlweb.pdf
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21(3), 354-375. <https://doi.org/10.1093/applin/21.3.354>
- Iwashita, N. (2001). The effect of learner proficiency on interactional moves and modified output in non-native-non-native interaction in Japanese as a foreign language. *System*, 29, 269-287. [https://doi.org/10.1016/S0346-251X\(01\)00015-X](https://doi.org/10.1016/S0346-251X(01)00015-X)
- Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, 29(1), 24-49. <https://doi.org/10.1093/applin/amm017>
- Japan Times. (2015, March 28). Disappointing levels of English. *Japan Times*. Retrieved July 10, 2017, from <http://www.japantimes.co.jp/opinion/2015/03/28/editorials/disappointing-levels-english/>
- Japan Times. (2016, February 15). Let's discuss English language education in Japan. *Japan Times*. Retrieved July 10, 2017, from <http://www.japantimes.co.jp/life/2016/02/15/language/lets-discuss-english-language-education-japan/>
- Jenkins, J. (2004). Research in teaching pronunciation and intonation. *Annual Review of Applied Linguistics*, 24, 109-125. <https://doi.org/10.1017/S0267190504000054>

- Kachru, B. B. (1985). Standards, codification, and sociolinguistic realism: The English language in the outer circle. In R. Quirk, & H. Widdowson (Eds.), *English in the World: Teaching and Learning the Language and the Literature*. Cambridge: Cambridge University Press.
- Kachru, B. B. (1986). World English: Agony and Ecstasy. *Journal of Aesthetic Education*, 30(2), 135-155. <https://doi.org/10.2307/3333196>
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387-412. <https://doi.org/10.1111/j.1467-1770.1990.tb00669.x>
- Levy, E., Leone, D., Moya-Gale, G., Hsu, S.-C., Chen, W., & Ramig, L. O. (2016). Vowel intelligibility in children with and without dysarthria. *Communication Disorders Quarterly*, 37(3), 171-179. <https://doi.org/10.1177/1525740115618917>
- Lindemann, S. (2002). Listening with an attitude: A model of native-speaker comprehension of non-native speakers in the United States. *Language in Society*, 31, 419-441. <https://doi.org/10.1017/S0047404502020286>
- Lindemann, S., & Subtirelu, N. (2012). Reliably biased: The role of listener expectation in the perception of second language speech. *Language Learning*, 63(3), 567-594. <https://doi.org/10.1111/lang.12014>
- Ministry of Education, Culture, Sports, Science and Technology, Japan (MEXT). (2014, January 23). *English Education Reform Plan Corresponding to Globalization*. Retrieved July 15, 2017, from http://www.mext.go.jp/en/news/topics/detail/_icsFiles/afieldfile/2014/01/23/1343591_1.pdf
- Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 49(1), 285-310. <https://doi.org/10.1111/0023-8333.49.s1.8>
- Nelson, C. L. (2008). Intelligibility since 1969. *World Englishes*, 27(3/4), 297-308. <https://doi.org/10.1111/j.1467-971X.2008.00568.x>
- Orikasa, M. (2016). The intelligibility of varieties of English in Japan. *World Englishes*, 35, 355-371. <https://doi.org/10.1111/weng.12209>
- Pennington, L. (1999). Assessing the communication skills of children with cerebral palsy: Does speech intelligibility make a difference? *Child Language Teaching and Therapy*, 15(2), 159-169. <https://doi.org/10.1177/026565909901500204>
- Perez, E. (1983). Disorders of speech intelligibility. *Journal of Childhood Communication Disorders*, 7(1), 59-75. <https://doi.org/10.1177/152574018300700107>
- Pickering, L. (2006). Current research on intelligibility in English as a lingua franca. *Annual Review of Applied Linguistics*, 26, 219-233. <https://doi.org/10.1017/S0267190506000110>
- Pickett, J. M., & Pollack, I. (1963). Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt. *Language and Speech*, 6(3), 151-164. <https://doi.org/10.1177/002383096300600304>
- Rajadurai, J. (2007). Intelligibility studies: A consideration of empirical and ideological issues. *World*

- Englishes*, 26(1), 87-98. <https://doi.org/10.1111/j.1467-971X.2007.00490.x>
- Rimmers, W. (2006). Measuring grammatical complexity: The Gordian knot. *Language Testing*, 23(4), 497-519. <https://doi.org/10.1191/0265532206lt339oa>
- Rivers, D. J. (2011). Intercultural processes in accented English. *World Englishes*, 30(3), 375-391. <https://doi.org/10.1111/j.1467-971X.2011.01707.x>
- Sewell, A. (2010). Research methods and intelligibility studies. *World Englishes*, 29(2), 257-269. <https://doi.org/10.1111/j.1467-971X.2010.01641.x>
- Skehan, P. & Foster, P. (2008). Complexity, accuracy, fluency and lexis in task-based performance: A meta-analysis of the Ealing research. In S. Van Daele, A. Housen, F. Kuiken, M. Pierrard, & I. Vedder (Eds.), *Complexity, Accuracy, and Fluency in Second Language Use, Learning, & Teaching* (pp. 207-226). Brussels: Contactforum.
- Tan, Y.-Y. (2015). "Native" and "non-native" perception of stress in Singapore English. *World Englishes*, 34(3), 355-369. <https://doi.org/10.1111/weng.12151>
- Thomson, R. I. (2012). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning*, 62(4), 1231-1258. <https://doi.org/10.1111/j.1467-9922.2012.00724.x>
- Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, 36(3), 326-344. <https://doi.org/10.1093/applin/amu076>
- Tran, M. (2015, November 17). Countries with high English Proficiency are more innovative. *Harvard Business Review*. Retrieved July 15, 2017, from <https://www.hbr.org/2015/11/countries-with-high-english-proficiency-are-more-innovative>
- Yuan, F., & Ellis, R. (2003). The effects of pre task planning and on line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, 24(1), 1-27. <https://doi.org/10.1093/applin/24.1.1>