

Trofimovich, P., Isaacs, T. Kennedy, S., & Tsunemoto, A. (in press). Speech comprehensibility. In T. M. Derwing, M. J. Munro, & R. I. Thomson, *The Routledge handbook of second language acquisition and L2 speaking*. Routledge.

Chapter 12

Speech Comprehensibility

Pavel Trofimovich

<https://orcid.org/0000-0001-6696-2411>

Talia Isaacs

<https://orcid.org/0000-0003-4302-3379>

Sara Kennedy

<https://orcid.org/0000-0001-5484-8048>

Aki Tsunemoto

<https://orcid.org/0000-0001-5916-8586>

In this chapter, we discuss the construct of second language (L2) comprehensibility, which refers to listeners' judgments of ease or difficulty in understanding L2 speech. We first define comprehensibility, highlighting terminological issues from a brief historical perspective, and distinguish this construct from related terms such as intelligibility. We review key evidence that supports the value of comprehensibility as a practical and useful measure of listeners' experience with L2 speech. We then describe current research findings suggesting that comprehensible L2 speech can be learned and taught, that L2 comprehensibility is linked to multiple linguistic and non-linguistic dimensions of speech, and that listeners' perceptions of speaker comprehensibility are dynamic and are influenced by social variables. We conclude by discussing ways in which

researchers, teachers, as well as speakers and listeners can approach L2 comprehensibility so as to capture its multifaceted nature.

INTRODUCTION

In 21st century second language (L2) pronunciation research, pedagogy, and assessment, two contrasting views continue to dominate the landscape (Levis, 2018). Propagated by the unregulated accent reduction industry (Thomson, 2013), the first view upholds nativelike attainment as the goal of L2 pronunciation learning and assessment. The second advances an agenda to help L2 speakers be more easily understandable, not necessarily nativelike, to listeners. This agenda includes researching the presumed factors that could foster or impede listeners' understanding of L2 speech and then developing pedagogical interventions that could enhance the quality of L2 speakers' performance and improve listeners' ability to understand them. Most L2 pronunciation experts deem the traditional emphasis on nativelikeness to be an unsuitable goal in many contexts of language use (Derwing & Munro, 2015). However, the alternative view—based on measuring how understandable L2 speakers are to listeners—has been mired in definitional confusion and inconsistency (Isaacs, 2008).

DEFINITIONS AND HISTORICAL PERSPECTIVE

An influential and widespread framework for listeners' understanding of L2 speech is Munro and Derwing's (1995a) distinction between two interrelated constructs: intelligibility and comprehensibility. Intelligibility, denoting the degree to which listeners actually understand L2 speech (see Chapter 11), has been operationalized in different ways, including the accuracy of listeners' orthographic transcriptions of L2 speech samples (Derwing & Munro, 1997; Munro, Derwing, & Morton, 2006) and listeners' responses to comprehension or true/false questions as a

check for understanding speech content (Hahn, 2004; Kennedy & Trofimovich, 2008). In contrast, comprehensibility denotes listeners' perceptions of the ease or difficulty with which they understand L2 speech. Comprehensibility is typically measured on a continuum, which is generally a rating scale. This construct has also been characterized by how irritating listeners perceive L2 speech to be, though this dimension is not emphasized in many studies (Ludwig, 1982; Piazza, 1980) likely because irritability (while related) does not fully overlap with comprehensibility.

Although Munro and Derwing's (1995a) distinction between intelligibility and comprehensibility is clear, it has not always been used in the same way, particularly in studies predating Munro and Derwing's influential work. Some researchers use the term intelligibility when measuring understanding through Likert-type scales (e.g., Fayer & Krasinski, 1987) when, in fact, what is being measured is comprehensibility. Other scholars use the term comprehensibility to refer to measures of what Munro and Derwing would call intelligibility, such as examining the accuracy of listeners' transcriptions of L2 utterances (e.g., Gass & Varonis, 1984, but see Varonis & Gass, 1982, for a measure compatible with Munro and Derwing's notion of comprehensibility). Yet others have used rating scales to measure intelligibility, which additionally conflate nativelike pronunciation and intelligibility. For example, Anderson-Hsieh, Johnson, and Koehler's (1992) scalar ratings of pronunciation ranged from "heavily accented speech that was unintelligible" to "near nativelike speech" (p. 538). This leads to two problems: listeners' perceptions are treated as their actual understanding, and speakers' comprehensibility is confounded with how nativelike they sound.

Definitional challenges can also be seen in the context of L2 oral proficiency scales in human-mediated standardized language tests often used for high-stakes decision making (e.g.,

TOEFL iBT, IELTS, TOEIC, Aptis), where the use of comprehensibility has become pervasive. Many rating scale descriptors make reference to intelligibility or intelligible speech, but the use of listener- or examiner-mediated scales implies that in fact, Munro and Derwing's notion of comprehensibility is being measured. To illustrate, Band 8 of the public version of the IELTS speaking descriptors refers to L2 speech as "easy to understand throughout; L1 [first language] accent has minimal effect on intelligibility" (British Council, 2020). In another example from language assessment, Isaacs, Trofimovich, and Foote (2018) developed a dedicated L2 comprehensibility scale with extended descriptors intended for English for Academic Purposes teachers to use as a pedagogical tool (i.e., for low-stakes formative assessment rather than high-stakes consequential decision making). In their detailed analytic scale, comprehensibility is discussed in terms of underlying pronunciation, fluency, lexis, and grammar features at different ability levels, with the degree of listener effort described across the subscales. This scale illustrates a data-driven approach to modeling comprehensibility, where comprehensibility is a multidimensional construct defined through multiple extended descriptors rather than a single numerical scale commonly used in research settings.

CRITICAL ISSUES

One overarching issue about the construct of comprehensibility is its role among other global measures of speaking (e.g., intelligibility) and specific metrics of speakers' performance (e.g., pronunciation accuracy, fluency). The key question is whether and to what degree scalar ratings of comprehensibility can be useful for language teachers and learners, researchers, and language speakers more generally.

Comprehensibility and Understanding

First and foremost, comprehensibility judgments can be useful to researchers and practitioners as a measure of L2 comprehension. Although intelligibility is often regarded as the gold standard for evaluating listeners' actual understanding of L2 speakers (Derwing & Munro, 2015), scalar ratings of comprehensibility are a useful measure in many contexts. To begin with, comprehensibility ratings are practical and intuitive, and can be elicited and scored easily using speech samples featuring the same content. In contrast, intelligibility measures require tasks with unique speech content for each instance when intelligibility is measured (to avoid greater intelligibility for content that is repeated to listeners) and comparatively more time for listeners to complete the tasks. Comprehensibility ratings are also reliable across listeners, meaning that listeners generally agree with each other regardless of how comprehensibility is measured (Munro, 2018; Nagle, 2019). By comparison, intelligibility scores often vary across task type, being influenced by the nature of the speech sample and the type of listening task used to measure intelligibility (Kang, Thomson, & Moran, 2018; Kennedy, 2009). Most importantly, although intelligibility and comprehensibility are partially independent, comprehensibility ratings provide a reasonable estimate of listeners' actual understanding of speech (Sheppard, Elliott, & Baese-Berk, 2017). For instance, Munro and Derwing (1995a) reported substantial overlaps between these dimensions, with correlation coefficients approaching .90, although the magnitude of this link might vary for different speakers and listeners (Matsuura, Chiba, & Fujieda, 1999). An intuitive, easy-to-use scalar measure, comprehensibility might thus be a useful general metric of understanding in a variety of contexts of language teaching, learning, and use.

Comprehensibility and Linguistic Content of Speech

Besides being a practical measure of understanding, comprehensibility ratings are also shaped by the linguistic content of speech, making them a useful metric of how linguistic dimensions of utterances impact the listener. By identifying linguistic correlates of comprehensibility, language researchers and teachers might target (through instruction or assessment) the dimensions of L2 speech that hinder listener processing. In their initial work, Munro and Derwing (1995a) found associations between listeners' comprehensibility ratings and several linguistic measures, including phonemic substitutions, intonation accuracy, and morphosyntactic errors. More recent work has revealed two constellations of linguistic dimensions relevant to comprehensibility: pronunciation (individual segments, prosody, fluency) and lexicogrammar (variety and richness of vocabulary, accuracy and complexity of grammar). The exact combinations of linguistic dimensions feeding into judgments of comprehensibility can depend on the speaker's linguistic background and the speaking task (Crowther, Trofimovich, Isaacs, & Saito, 2018), but the general finding has been consistent. Many measures at the level of segments, prosody, fluency, grammar, and discourse have been linked to listeners' ratings of comprehensibility in multiple languages (Isaacs & Trofimovich, 2012; Saito, Trofimovich, & Isaacs, 2017a). To take an example from L2 English, Kang, Rubin, and Pickering (2010) reported that 50% of the variance in comprehensibility can be explained by fluency and prosody (e.g., intonation, pausing, speech rate). For L2 French, Bergeron and Trofimovich (2017) found links between comprehensibility ratings and measures of pronunciation, lexicogrammar, fluency, and discourse richness. In L2 German, O'Brien (2014) showed that comprehensibility was tied to fluency and accuracy measures for vocabulary, morphology, and pronunciation. The assumption underlying this work is that teachers might specifically target these linguistic dimensions through instruction to help

L2 speakers become more comprehensible to interlocutors.

Comprehensibility and Processing Fluency

Comprehensibility might also be useful to researchers and practitioners as a measure of processing fluency, defined as a listener's subjective experience of the ease or difficulty with which information is processed (Reber & Greifeneder, 2017; Schwarz, 2018). A key aspect of processing fluency which cuts across various social and psychological domains is that people appraise and respond to various situations based on the perceived difficulty they report while processing a stimulus (e.g., text, image, sound), which may or may not reflect their actual experience with that stimulus. For instance, statements attributed to people whose names are harder to pronounce are considered less trustworthy (Newman et al., 2014), regardless of the content of the statements. Similarly, readers exposed to text printed in a difficult-to-read font react more negatively than those reading the same text in an easy-to-read font, despite having similar text comprehension for both conditions (Sanchez & Jaeger, 2015; Song & Schwarz, 2008). Munro and Derwing (1995a) observed that comprehensibility might be rated differently for speech that is perfectly intelligible, which aligns with findings from processing fluency studies that listeners' various reactions to speech and speakers might be linked not to actual understanding (intelligibility) but to comprehensibility.

Growing evidence suggests that comprehensibility captures socially important decisions for listeners. For instance, in social-psychological research, speakers who listeners perceived as hard to understand were downgraded in listeners' affective and attitudinal evaluations. Such speakers were ascribed negative emotions of annoyance and irritation and deemed less intelligent and successful (Dragojevic, Giles, Beck, & Tatum, 2017). Similarly, in an e-learning study, when students evaluated an instructional video narrated by the instructor who was rated hard to

understand, students downgraded their evaluations of the instructor, expressed negative attitudes towards coursework, and evaluated video content as more difficult, even though students' actual understanding of the video was not compromised (Sanchez & Khan, 2016). In fact, a comprehensibility scale akin to that used in L2 speech research has now been validated as part of a five-item processing fluency measure that appears to explain various human judgments (truthfulness, preference, perceived risk) all formerly attributed to processing fluency (Graf, Mayer, & Landwehr, 2018). Thus, an intuitive appeal of comprehensibility as a measure of processing fluency is that it might help explain aspects of human behavior, including, for instance, whether interlocutors continue interacting with speakers they find difficult to understand or whether university students drop out of courses led by instructors whose speech they consider hard to process.

Comprehensibility and the (Imagined) Interlocutor

The usefulness of comprehensibility as a measure of L2 speakers' success at conveying their message or as a measure of the linguistic dimensions that matter most for comprehensibility hinges on the key issue of who is judging comprehensibility. If the response to this question is native speakers, then the follow-up question should be, Why are native speakers the only suitable judges of comprehensibility?

Whereas early L2 comprehensibility research relied nearly exclusively on native-speaking listeners (e.g., Munro & Derwing, 1995a), more recently, researchers have employed more varied listener groups, including L2 speakers (Crowther, Trofimovich, & Isaacs, 2016; Derwing & Munro, 2013; O'Brien, 2014), bilinguals and multilinguals (Saito & Shintani, 2016), and members of specific academic and professional groups regardless of native speaker status (Derwing & Munro, 2009; Kennedy & Trofimovich, 2013; Sheppard et al., 2017). Prior research

has revealed multiple listener characteristics that impact comprehensibility judgments, including listeners' experience with the language being evaluated (Munro et al., 2006), listeners' teaching experience (Saito, Trofimovich, Isaacs, & Webb, 2017b) and linguistic training (Isaacs & Thomson, 2013), listener status as bilinguals or multilinguals (Saito & Shintani, 2016), and their L2 learning experience (Saito et al., 2019). Nevertheless, despite minor differences, various listener groups appear to be similar in the quality and consistency of the comprehensibility ratings they assign to the same L2 speakers (Crowther et al., 2016; Derwing & Munro, 2013; Saito et al., 2017b), even though listeners might rely on different criteria to arrive at similar ratings (Foote & Trofimovich, 2018; Isaacs & Thomson, 2013). In light of this evidence, although researchers may find it practical to elicit only "native" listeners' comprehensibility ratings, this would seem short-sighted. It may be irresponsible to presume that L2 speakers will exclusively speak with native speakers, especially for languages of major global or regional significance (e.g., English, Mandarin, Spanish).

CURRENT CONTRIBUTIONS AND RESEARCH

Although early research focusing on comprehensibility predominantly examined speaker and listener characteristics relevant to comprehensibility (e.g., Munro & Derwing, 1995a; Gass & Varonis, 1984; Tyler & Bro, 1992), more recent work has adopted a multidimensional perspective, exploring comprehensibility as a socially driven, dynamic construct that is interdependent on both speaker and listener. Researchers have also intensified applied research examining the effect of instructional interventions on comprehensibility.

Comprehensibility – Pedagogically Relevant

One rapidly expanding and valuable research strand focuses on how to help L2 speakers to speak comprehensibly. Comprehensible speech takes time and effort to develop. L2 speakers are not

generally judged more comprehensible when performing the same speaking task a second time; nor do they sound any more comprehensible to listeners after being told to make their speech as easy for the interlocutor to understand as possible (Strachan, Trofimovich, & Kennedy, 2019). Similarly, comprehensibility might not greatly improve for L2 speakers by virtue of attending a university in the target language, in the absence of pronunciation instruction (Kennedy, Foote, & Buss, 2015). Even when instruction is available, improving comprehensibility is not guaranteed during one academic term (Kennedy & Trofimovich, 2010). In a landmark 7-year study, Derwing and Munro (2013) showed that Slavic but not Mandarin adult immigrants to Canada showed an improvement in their L2 English comprehensibility ratings across three timepoints (2 months, 2 years, and 7 years of residence). The success of the Slavic group was attributed to their greater integration into English-speaking communities and greater exposure to and use of English. Thus, even after many years of language use opportunities, some L2 speakers may not adopt speech patterns that would be judged easier to understand, suggesting that comprehensibility is an important target for instruction.

Fortunately, comprehensibility can be improved through instruction. For instance, Derwing, Munro, and Wiebe (1998) showed that supplementing regular language instruction with focused teaching on speech fluency (e.g., speaking rate) and prosody (e.g., intonation, stress, rhythm) led to significant improvement in L2 learners' comprehensibility after 12 weeks. Extensive interaction through video-conferencing (Saito & Akiyama, 2017) and specific teaching techniques such as shadowing (Foote & McDonough, 2017) have also been shown to positively impact comprehensibility. In a rare longitudinal study of comprehensibility development during one year of university instruction in L2 Spanish, Nagle (2018) showed that learners' comprehensibility improved through communicative teaching alone, without a dedicated

pronunciation focus, and that the extent of improvement was unrelated to learners' motivational profiles. In contrast, Saito, Dewaele, Abe, and In'nami (2018) reported links between Japanese high school students' comprehensibility gains and measures capturing students' motivation and positive feelings towards their learning. It appears that focused instruction is useful even for learners who have become entrenched in their speech patterns. Derwing, Munro, Foote, Waugh, and Fleming (2014) taught a 17-hour pronunciation course to L2-speaking employees at a Canadian factory. These L2 speakers, who on average had resided in Canada for 19 years, showed significant progress in their comprehensibility ratings after instruction. An interim conclusion is that comprehensible L2 speech can be taught and learned, which has prompted researchers to promote comprehensibility as a measure of L2 learning compatible with principles of communicative language teaching (Saito & Plonsky, 2019).

Comprehensibility – Dynamic

Although pedagogically-oriented investigations of comprehensibility have tracked L2 speakers' comprehensibility over weeks, months, and sometimes years, comprehensibility has rarely been framed as a dynamic, variable process which can change on a finer-grained timescale, as a matter of minutes or seconds. Nagle, Trofimovich, and Bergeron (2019) explored whether comprehensibility can be construed as dynamic, examining how raters explain their assessments as they evolve over time. Twenty-four Spanish-speaking listeners evaluated 3-minute personal narratives by L2 Spanish learners using a computer interface which allowed listeners to increase or decrease the comprehensibility rating as the speech unfolded. Listeners showed varying rating profiles, such that some listeners increased or decreased comprehensibility ratings infrequently over a speech sample whereas others increased or decreased ratings at a high frequency, with varying magnitude of change. In a follow-up study, Trofimovich et al. (2020) reasoned that

interactive speech, where interlocutors react to one another in real time, might be even more amenable to dynamic comprehensibility judgements, compared to the one-way listening task that Nagle et al. (2019) examined. For this study, L2 English university students from different language backgrounds engaged in collaborative tasks over 17 minutes, rating their partner's comprehensibility at 2–3 minute intervals. Speakers' comprehensibility ratings for the most part followed a U-shaped function, with comprehensibility (initially perceived to be high) dipping to lower levels but then reaching high levels by the end of the interaction. Speakers' ratings also became more similar to each other soon after the interaction started and remained alike throughout. Taken together, these findings not only suggest that speakers' comprehensibility can change over time as interaction unfolds but also imply that comprehensibility issues might become less important for both interlocutors in a conversation after a certain minimum threshold of comprehensibility has been reached. Whether such a threshold involves a degree of interpersonal comfort or is simply a matter of investing sufficient time into communication is an area for future research.

Comprehensibility – Socially Flexible

That listeners' understanding of speech is subject to social influences is relatively well known. For instance, Rubin and Smith (1990) showed that perceived speaker ethnicity influenced listeners' comprehension of a lecture. Listeners who were led to believe that the speaker was Asian demonstrated less understanding than those who were told that the same speaker was Caucasian. Recent work has revealed several similar social influences on comprehensibility. For instance, Sheppard et al. (2017) showed that university faculty who reported negative attitudes towards the English proficiency of international students gave lower comprehensibility ratings to students' L2 speech than did faculty with positive attitudes, despite both groups transcribing the

speech with equal accuracy. In another study, Taylor Reid, Trofimovich, and O'Brien (2019) presented listeners with either a positive or a negative biasing anecdote about language abilities of L2 speakers, before asking listeners to evaluate comprehensibility. Compared to the assessment of baseline listeners (no anecdote), positively-oriented listeners rated speakers more favorably while negatively-oriented listeners (especially older individuals with likely more entrenched social attitudes) evaluated the same speakers more negatively. Taylor Reid, O'Brien, Trofimovich, and Bajt (2020) showed that teachers of L2 German, particularly native speakers of German, were similarly influenced by negative comments, downgrading L2 German learners' ratings relative to the assessment of baseline listeners. However, it appears that positive and negative social biases can be mitigated through relatively simple interventions. In a follow-up study, Taylor Reid, Trofimovich, and O'Brien (in review) demonstrated that perspective taking (Hansen, Rakic, & Steffens, 2014; Weyant, 2007), which they defined as asking listeners to practice the target speech task before engaging in speech rating, was effective at eliminating negative bias effects in rating comprehensibility (see also Taylor Reid, O'Brien, Trofimovich, & Tsunemoto, in press). Put differently, encouraging listeners to "walk in the shoes" of people whose speech was evaluated may have immunized listeners against various attitudinal comments that they were exposed to before the rating. These findings underscore comprehensibility as a socially and contextually flexible construct influenced by positive and negative biases and highlight the value of interventions that can minimize these biases in research and practice settings.

MAIN RESEARCH METHODS

In research contexts, comprehensibility is most often measured using 9-point numerical rating scales (Munro & Derwing, 1995a), although other scale lengths have been attested, including 5-

point scales (Isaacs & Thomson, 2013), 7-point scales (Ludwig & Mora, 2017), 8-point scales (Polyanskaya & Ordin, 2019), and 10-point scales (Caspers, 2010), often without thorough validation (Thomson, 2018). In Munro and Derwing's (1995a) initial study, the lowest scalar value is described as "extremely easy to understand," whereas the highest scalar value designates speech that is "impossible to understand" (p. 79). Other researchers reverse the scalar extremes to approximate what they consider a more intuitive scale direction, where a higher value represents better (more comprehensible) performance (Isaacs & Trofimovich, 2012), or replace verbal endpoint descriptions with images of smiling and frowning faces illustrating scale directionality (Taylor Reid et al., 2019).

Researchers have occasionally opted for continuous scales over Likert scales. In a paper-and-pencil format, raters indicate their judgment by marking a location on a straight line bounded by endpoint descriptors, and researchers measure the distance (e.g., in millimeters) between the left endpoint and the mark (Isaacs, Trofimovich, Yu, & Chereau, 2015). In a computer-based format, raters move a slider to record their rating, typically on a 1,000-point continuum (Saito et al., 2017a). Another approach is to ask raters to estimate the proportion of words that they can understand using a percentage scale (Isaacs, 2008; Kang, Thomson, & Moran, 2018). Yet another option is to measure comprehensibility through direct magnitude estimation by comparing the target speech sample with a reference item (Munro, 2018). As an index of processing difficulty, comprehensibility has also been operationalized in some research in terms of the time it takes for listeners to process speech content (Ludwig & Mora, 2017; Munro & Derwing, 1995b) or examined through listeners' performance in a concurrent reaction-time task, such as monitoring tones while trying to understand and remember the lecture's content (Hahn, 2004). More recently, comprehensibility has been measured dynamically, using

Idiodynamic Software (MacIntyre, 2012), with listeners upgrading and downgrading ratings in real time as they experience L2 speech (Nagle et al., 2019).

Most comprehensibility studies have targeted English, with few studies focusing on other languages, including German (O'Brien, 2014), French (Bergeron & Trofimovich, 2017), Spanish (Nagle, 2018), Japanese (Saito & Akiyama, 2017), and Korean (Isbell, Park, & Lee, 2019).

Target speakers are mostly university students but rarely school-age learners (e.g., Saito et al., 2018) and almost never children or older individuals (but see Derwing et al., 2014). To elicit speech for comprehensibility rating, researchers have used controlled read-aloud tasks involving individual words (e.g., Caspers, 2010) and sentences (e.g., Kennedy & Trofimovich, 2008), as well as various tasks eliciting monologic, extemporaneous speech performances, including timed and untimed picture descriptions (e.g., Derwing, Munro, Thomson, 2008), tasks from operational L2 speaking tests (e.g., Isaacs et al., 2015) or practice test items (e.g., Crowther et al., 2018), argumentative tasks (e.g., Suzuki & Kormos, 2019), and mock job interviews (e.g., Kennedy & Trofimovich, 2013). Only recently have researchers begun investigating comprehensibility for both members of dyads engaged in conversation (e.g., Trofimovich et al., 2020).

Comprehensibility ratings are often collected through paper-and-pencil questionnaires but increasingly so through online interfaces (e.g., Crowther et al., 2016; Saito & Shintani, 2016) and online crowdsourcing platforms (Nagle, 2019).

Several generalizations have emerged from past methodologically-oriented research focusing on speech ratings:

- There is little difference in the ratings obtained through the use of 5- versus 9-point scales, although shorter scales are sometimes perceived by listeners as constraining whereas longer scales are considered difficult for differentiating across skill levels

(Isaacs & Thomson, 2013). Compared to direct magnitude estimation of comprehensibility, 9-point scales perform just as well, suggesting that the use of scalar ratings is a reliable approach to measuring L2 comprehensibility for research purposes (Munro, 2018).

- Evaluations of individual short sentences by the same speaker often lack consistency, suggesting that ratings of shorter speech samples might not be representative of ratings of longer discourse produced by the same speaker (Munro, 2018).
- Listeners sometimes assign harsher ratings when evaluating the same samples again, because listeners might become increasingly aware of how the speakers' output differs from the language expected by listeners (Flege & Fletcher, 1992; Munro & Derwing, 1994).
- Comprehensibility ratings do not appear to be influenced by whether this dimension is evaluated separately or in combination with other global dimensions such as accentedness and fluency (O'Brien, 2016) or by the order in which comprehensibility judgments occur in a rating sequence (Derwing & Munro, 1997; O'Brien, 2016).
- Speech ratings obtained in online environments with built-in controls (e.g., through crowdsourcing platforms) yield highly reliable judgments, comparable to those obtained in research laboratories (Nagle, 2019).

As Isaacs et al. (2015) note, regardless of the method used to capture comprehensibility, in the absence of detailed guidance, raters may interpret the target construct in different ways, for example, assuming that it refers to listeners' perceptions of understanding the overall message, to understanding every single word that is uttered, or solely to understanding meaning-laden words. Put simply, some listeners' interpretations of comprehensibility may differ from other listeners'

interpretations and might not exactly conform to what the researchers believe they are measuring. This is important to establish in light of construct validity for comprehensibility measurement, which has only been examined infrequently to date (Isaacs & Thomson, 2013; Munro, 2018; Nagle, 2019).

RECOMMENDATIONS FOR PRACTICE

With respect to implications of comprehensibility research for language teaching, it is encouraging that many researchers and teachers see comprehensibility as a construct with multiple elements, not just pronunciation. Elements that may be relevant or important should be highlighted as much as possible in teaching and learning materials and in contexts for teaching or tutoring L2 speech or familiarizing listeners with L2 speech. For instance, the type of speaking task, the speaker's use of vocabulary and grammar, the listener's level of motivation, attitude towards or experience with L2 speech and learning generally are all elements that could be linked to comprehensibility. Teachers, speakers, and listeners could then work on elements over which they potentially have control, such as pronunciation, vocabulary, attitude, motivation, or experience with L2 speech. Clearly, neither teachers and speakers nor listeners can attend simultaneously to all elements potentially linked to comprehensibility. The importance of particular elements varies according to the person and context, and speakers and listeners should be encouraged to enhance their awareness of these elements through awareness-building activities, including guided analysis of self or others' comprehensibility or spoken language (e.g., Derwing, Rossiter, & Munro, 2002; Krech Thomas, 2004).

In light of comprehensibility's dynamic nature, it is important for researchers to consider how the length of a speech sample or frequency of rating might impact comprehensibility. Because comprehensibility trends upward over the course of interactions, L2 speakers aiming for

comprehensibility or increased confidence in their comprehensibility should be encouraged to seek opportunities for spoken interactions which are not brief. These might be found in group discussions or brainstorming sessions, interviews, workshops, and community group meetings. Another consideration is speakers' overall affective and motivational profiles which have been linked to comprehensibility ratings. Many researchers and teachers try to ensure that their research and teaching contexts are not stressful for speakers or listeners. Rehearsal of tasks and self-reports of anxiety, willingness to communicate, and motivation could also help to modulate or document the possible influence of these and other variables on interlocutors' comprehensibility. Confidence could also be promoted by teachers or L2 speakers through calming or self-affirming exercises prior to and during spoken interaction. On a more global scale, enjoyable learning environments, where teachers promote positivity, encourage learners' desire to communicate, increase motivation, and reduce anxiety (e.g., Moskowitz & Dewaele, 2019), are likely most conducive to the development of comprehensibility and successful L2 learning generally.

Because comprehensibility appears to be influenced by social variables, language teachers might engage their learners in initiatives which involve structured opportunities for positive contact between various types of interlocutors. Other initiatives might target native-speaking listeners, to help them discover some differences between their language and another language, do structured practice in transcribing L2 speech, or take the perspective of an L2 peer. The goal would be to guide people to consider different facets of individuals with whom they might unknowingly share linguistic and social commonalities, as a way of promoting harmonious communication (Hansen et al., 2014). With encouragement and support from administration and managers, formal or informal activities such as happy hours, sharing circles, or language classes

can also be done in workplaces with colleagues from different backgrounds (Kim, Roberson, Russo, & Briganti, 2019). Of course, individuals can themselves initiate contact with L2 speakers or try to learn or use a less familiar language, and so reduce anxiety or develop more positive attitudes about communication with L2 speakers. For research, the importance of attitudes towards L2 speech means that eliciting measures of raters' attitudes will add another dimension to the analysis and interpretation of comprehensibility ratings. Moreover, social biases which a rater is exposed to before rating can affect the rating itself. Researchers should carefully consider who is involved in administering the rating session (e.g., a majority or minority language speaker in a given context), what is said prior to the ratings, and how these factors could influence raters and the scores they assign.

FUTURE DIRECTIONS

Comprehensibility is an appealing construct because it connects language learners and teachers, who might be interested in improving L2 oral production, with researchers, whose goal is to describe what linguistic, social, experiential, and behavioural dimensions underlie people's experience with speech. The breadth of theoretical questions and the versatility of applied contexts relevant to comprehensibility make for exciting future research. For example, researchers could intensify longitudinal research examining how learners with different cognitive, motivational, experiential, and affective profiles develop comprehensible L2 speech across different contexts, both instructed and uninstructed. In keeping with a dynamic view of comprehensibility, researchers might continue exploring interlocutors' comprehensibility in paired or group interaction. This work could clarify how interlocutors' cumulative shared experience impacts their comprehensibility ratings in tasks that increase versus decrease in cognitive difficulty over time and examine how non-linguistic cues (e.g., gestures, facial

expressions, displays of emotion) and interactional variables (e.g., backchanneling, clarification requests) contribute to interlocutors' mutual comprehensibility judgments.

Researchers might also explore links between interaction-based comprehensibility ratings and interlocutor awareness of what makes speech comprehensible for them, using different combinations of interlocutors who vary in language proficiency, experience, and other variables (e.g., personality characteristics). Similarly, it might be useful to explore long-term effects of interlocutors' extended conversational experience on their perception of comprehensibility, focusing on speakers' judgments of the same and new partners in another instance of interaction, after a delay. In light of demonstrated alignment between both partners' comprehensibility scores in extended interaction, it could also be fruitful to examine the validity of a joint (rather than speaker-specific) measure of comprehensibility for both partners in a conversational dyad. Given attitudinal influences on comprehensibility, researchers might explore situations where interlocutors' comprehensibility judgments are influenced by one or both interlocutors' sociopolitical views, stereotypical judgments, or other attitudes towards the speaker or the topic of conversation. Finally, comprehensibility ratings, as useful measures of listener understanding and listener processing fluency, could be examined in relation to such conversational phenomena as speakers' engagement in dialogue, participation patterns, or affective responses to the task or their partner, to clarify the role of processing effort in interlocutor experience in interaction.

CONCLUSION

Beyond a doubt, comprehensibility is a valuable construct, relevant to both speakers and listeners and useful for both researchers and educational practitioners. As discussed in this chapter, people's perceptions of each other's comprehensibility are subject to social influences, evolve dynamically over time as communication unfolds, are tied to many linguistic (and some non-

linguistic) features of interaction, and affect other aspects of people's judgments, such as how annoying or intelligent a person is. These characteristics make comprehensibility a worthy conceptual and practical target. By understanding how interlocutors perceive each other's comprehensibility (in terms of what comprehensibility means for them), it might be possible to empower speakers to become more successful L2 communicators. Nevertheless, comprehensibility is but one of several possible constructs relevant to L2 speech. To gain a clearer understanding of the teaching and learning of L2 speech would unquestionably require the use of multiple complementary metrics of listeners' understanding, including measures of comprehensibility, intelligibility, and listening comprehension.

FURTHER READING

- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam: John Benjamins. This book features a state-of-the-art review of literature on various constructs relevant to L2 pronunciation, including comprehensibility.
- Isaacs, T., & Trofimovich, P. (Eds.). (2016). *Second language pronunciation assessment: Interdisciplinary perspectives*. Bristol, UK: Multilingual Matters. This open access edited volume features multiple research contributions relevant to pronunciation assessment, including the assessment of comprehensibility.
- Isaacs, T., Trofimovich, P., & Foote, J. A. (2018). Developing a user-oriented second language comprehensibility scale for English-medium universities. *Language Testing*, 35, 193–216. This study describes the development and validation of a comprehensibility-focused scale for English for Academic Purposes teachers.

Nagle, C., Trofimovich, P., & Bergeron, A. (2019). Toward a dynamic view of second language comprehensibility. *Studies in Second Language Acquisition*, *41*, 647–672. This is the first study investigating comprehensibility from a dynamic perspective.

REFERENCES

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, *42*, 529–555.

Bergeron, A., & Trofimovich, P. (2017). Linguistic dimensions of accentuatedness and comprehensibility: Exploring task and listener effects in second language French. *Foreign Language Annals*, *50*, 547–566.

British Council (2020). *How IELTS is assessed*. Retrieved 5 April 2020 from <https://takeielts.britishcouncil.org/teach-ielts/test-information/assessment>

Caspers, J. (2010). The influence of erroneous stress position and segmental errors on intelligibility, comprehensibility and foreign accent in Dutch as a second language. *Linguistics in the Netherlands*, *27*, 17–29.

Crowther, D., Trofimovich, P., & Isaacs, T. (2016). Linguistic dimensions of second language accent and comprehensibility: Nonnative listeners' perspectives. *Journal of Second Language Pronunciation*, *2*, 160–182.

Crowther, D., Trofimovich, P., Saito, K., & Isaacs, T. (2018). Linguistic dimensions of L2 accentuatedness and comprehensibility vary across speaking tasks. *Studies in Second Language Acquisition*, *40*, 443–457.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, *19*, 1–16.

- Derwing, T. M., & Munro, M. J. (2009). Comprehensibility as a factor in listener interaction preferences: Implications for the workplace. *Canadian Modern Language Review*, 66, 181–202.
- Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, 63, 163–185.
- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam: John Benjamins.
- Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, 29, 359–380.
- Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, 48, 393–410.
- Derwing, T. M., Munro, M. J., Foote, J. A., Waugh, E., & Fleming, J. (2014). Opening the window on comprehensible pronunciation after 19 years: A workplace training study. *Language Learning*, 64, 526–548.
- Derwing, T. M., Rossiter, M. J., & Munro, M. J. (2002). Teaching native speakers to listen to foreign-accented speech. *Journal of Multilingual and Multicultural Development*, 23, 245–259.
- Dragojevic, M., Giles, H., Beck, A.-C., & Tatum, N. T. (2017). The fluency principle: Why foreign accent strength negatively biases language attitudes. *Communication Monographs*, 84, 385–405.
- Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning*, 37, 313–326.
- Flege, J., & Fletcher, K. (1992). Talker and listener effects on the perception of degree of foreign

- accent. *Journal of the Acoustical Society of America*, 91, 370–389.
- Foote, J. A., & McDonough, K. (2017). Using shadowing with mobile technology to improve L2 pronunciation. *Journal of Second Language Pronunciation*, 3, 34–56.
- Foote, J., & Trofimovich, P. (2018). Is it because of my language background? A study of language background influence on comprehensibility judgments. *Canadian Modern Language Review*, 74, 253–278.1
- Gass, S., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning*, 34, 65–89.
- Graf, L. K. M., Mayer, S., & Landwehr, J. R. (2018). Measuring processing fluency: One versus five items. *Journal of Consumer Psychology*, 28, 393–411.
- Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38, 201–223.
- Hansen, K., Rakic, T., & Steffens, M. C. (2014). When actions speak louder than words: Preventing discrimination of nonstandard speakers. *Journal of Language and Social Psychology*, 33, 68–77.
- Isaacs, T. (2008). Towards defining a valid assessment criterion of pronunciation proficiency in non-native English-speaking graduate students. *Canadian Modern Language Review*, 64, 555–580.
- Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, 10, 135–159.
- Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second*

- Language Acquisition*, 34, 475–505.
- Isaacs, T., Trofimovich, P., & Foote, J. A. (2018). Developing a user-oriented second language comprehensibility scale for English-medium universities. *Language Testing*, 35, 193–216.
- Isaacs, T., Trofimovich, P., Yu, G., & Chereau, B. M. (2015). Examining the linguistic aspects of speech that most efficiently discriminate between upper levels of the revised IELTS Pronunciation scale. *IELTS Research Report Series*, 4, 1–48.
- Isbell, D. R., Park, O. S., & Lee, K. (2019). Learning Korean pronunciation: Effects of instruction, proficiency, and L1. *Journal of Second Language Pronunciation*, 5, 13–48.
- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38, 301–315.
- Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, 94, 554–566.
- Kang, O., Thomson, R. I., & Moran, M. (2018). Empirical approaches to measuring the intelligibility of different varieties of English in predicting listener comprehension. *Language Learning*, 68, 115–146.
- Kennedy, S. (2009). L2 proficiency: Measuring the intelligibility of words and extended speech. In A. Benati (Ed.), *Issues in second language proficiency* (pp. 132–144). London: Continuum.
- Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review*, 64, 459–490.

- Kennedy, S., & Trofimovich, P. (2010). Language awareness and second language pronunciation: A classroom study. *Language Awareness, 19*, 171–185.
- Kennedy, S., & Trofimovich, P. (2013). First-and final-semester non-native students in an English-medium university: Judgments of their speech by university peers. *Learning in Higher Education, 3*, 283–303.
- Kennedy, S., Foote, J. A., & Buss, L. K. (2015). Second language speakers at university: Longitudinal development and rater behaviour. *TESOL Quarterly, 49*, 199–209.
- Kim, R., Roberson, L., Russo, M., & Briganti, P. (2109). Language diversity, nonnative accents, and their consequences at the workplace: Recommendations for individuals, teams, and organizations. *The Journal of Applied Behavioral Science, 55*, 73–95.
- Krech Thomas, H. (2004). *Training strategies for improving listeners' comprehension of foreign-accented speech* (Doctoral dissertation). Retrieved from https://scholar.colorado.edu/concern/graduate_thesis_or_dissertations/j098zb31g
- Levis, J. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge: Cambridge University Press.
- Ludwig, A., & Mora, J. C. (2017). Processing time and comprehensibility judgments in non-native listeners' perception of L2 speech. *Journal of Second Language Pronunciation, 3*, 167–198.
- Ludwig, J. (1982). Native-speaker judgments of second-language learners' efforts at communication: A review. *The Modern Language Journal, 66*, 274–283.
- MacIntyre, P. D. (2012). The idiodynamic method: A closer look at the dynamics of communication traits. *Communication Research Reports, 29*, 361–367.

- Matsuura, H., Chiba, R., & Fujieda, M. (1999). Intelligibility and comprehensibility of American and Irish Englishes in Japan. *World Englishes*, 18, 49–62.
- Moskowitz, S., & Dewaele, J.-M. (2019). Is teacher happiness contagious? A study of the link between perceptions of language teacher happiness and student attitudes. *Innovation in Language Learning and Teaching*. Advanced Online Publication.
- Munro, M. J. (2018). Dimensions of pronunciation. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation* (pp. 413–431). New York: Routledge.
- Munro, M. J., & Derwing, T. M. (1994). Evaluations of foreign accent in extemporaneous and read material. *Language Testing*, 11, 253–266.
- Munro, M. J., & Derwing, T. M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45, 73–97.
- Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, 38, 289–306.
- Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28, 113–131.
- Nagle, C. (2018). Motivation, comprehensibility, and accentedness in L2 Spanish: Investigating motivation as a time-varying predictor of pronunciation development. *The Modern Language Journal*, 102, 199–217.
- Nagle, C. (2019). Developing and validating a methodology for crowdsourcing L2 speech ratings in Amazon Mechanical Turk. *Journal of Second Language Pronunciation*, 5, 294–323.
- Nagle, C., Trofimovich, P., & Bergeron, A. (2019). Toward a dynamic view of second language comprehensibility. *Studies in Second Language Acquisition*, 41, 647–672.

- Newman, E. J., Sanson, M., Miller, E. K., Quigley-McBride, A., Foster J. L., Bernstein, D. M., & Garry, M. (2014). People with easier to pronounce names promote truthiness of claims. *PLoS ONE*, *9*(2), e88671.
- O'Brien, M. G. (2014). L2 learners' assessments of accentedness, fluency, and comprehensibility of native and nonnative German speech. *Language Learning*, *64*, 715–748.
- O'Brien, M. G. (2016). Methodological choices in rating speech samples. *Studies in Second Language Acquisition*, *38*, 587–605.
- Piazza, L. G. (1980). French tolerance for grammatical errors made by Americans. *The Modern Language Journal*, *64*, 422–427.
- Polyanskaya, L., & Ordin, M. (2019). The effect of speech rhythm and speaking rate on assessment of pronunciation in a second language. *Applied Psycholinguistics*, *40*, 795–819.
- Reber, R., & Greifeneder, R. (2017) Processing fluency in education: How metacognitive feelings shape learning, belief formation, and affect. *Educational Psychologist*, *52*, 84–103.
- Rubin, D. L., & Smith, K. A. (1990). Effects of accent, ethnicity, and lecture topic on undergraduates' perceptions of nonnative English-speaking teaching assistants. *International Journal of Intercultural Relations*, *14*, 337–353.
- Saito, K., & Akiyama, Y. (2017). Linguistic correlates of comprehensibility in second language Japanese speech. *Journal of Second Language Pronunciation*, *3*, 199–217.
- Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, *69*, 652–708.

- Saito, K., & Shintani, N. (2016). Do native speakers of North American and Singapore English differentially perceive comprehensibility in second language speech? *TESOL Quarterly*, *50*, 421–446.
- Saito, K., Dewaele, J.-M., Abe, M., & In'nami, Y. (2018). Motivation, emotion, learning experience, and second language comprehensibility development in classroom settings: A cross-sectional and longitudinal study. *Language Learning*, *68*, 709–743.
- Saito, K., Tran, M., Suzukida, Y., Sun, H., Magne, V., & Ilkan, M. (2019). How do L2 listeners perceive the comprehensibility of foreign-accented speech? Roles of L1 profiles, L2 proficiency, age, experience, familiarity and metacognition. *Studies in Second Language Acquisition*, *41*, 1133–1149.
- Saito, K., Trofimovich, P., & Isaacs, T. (2017a). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*, 439–462.
- Saito, K., Trofimovich, P., Isaacs, T., & Webb, S. (2017b). Re-examining phonological and lexical correlates of second language comprehensibility: The role of rater experience. In T. Isaacs & P. Trofimovich (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (pp. 141–156). Bristol, UK: Multilingual Matters.
- Sanchez, C. A., & Jaeger, A. J. (2015). If it's hard to read, it changes how long you do it: Reading time as an explanation for perceptual fluency effects on judgment. *Psychonomic Bulletin and Review*, *22*, 206–211.
- Sanchez, C. A., & Khan, S. (2016). Instructor accents in online education and their effect on learning and attitudes. *Journal of Computer Assisted Learning*, *32*, 494–502.
- Schwarz, N. (2018). Of fluency, beauty, and truth: Inferences from metacognitive experiences. In

- J. Proust & M. Fortier (Eds.), *Metacognitive diversity: An interdisciplinary approach* (p. 25–46). Oxford, UK: Oxford University Press.
- Sheppard, B. E., Elliott, N. C., & Baese-Berk, M. M. (2017). Comprehensibility and intelligibility of international student speech: Comparing perceptions of university EAP instructors and content faculty. *Journal of English for Academic Purposes, 26*, 42–51.
- Song, H., & Schwarz, N. (2008). If it's hard to read, it's hard to do: Processing fluency affects effort prediction and motivation. *Psychological Science, 19*, 986–988.
- Strachan, L., Kennedy, S., & Trofimovich, P. (2019). Second language speakers' awareness of their own comprehensibility: Examining task repetition and self-assessment. *Journal of Second Language Pronunciation, 5*, 347–373.
- Suzuki, S., & Kormos, J. (2020). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition, 42*, 143–167.
- Taylor Reid, K., O'Brien, M. G., Trofimovich, P., & Tsunemoto, A. (in press). Exploring the stability of second language speech ratings through task practice in bilinguals' two languages. *Journal of Monolingual and Bilingual Speech*.
- Taylor Reid, K., O'Brien, M., Trofimovich, P., & Bajt, A. (2020). Testing the malleability of teachers' judgments of second language speech. *Journal of Second Language Pronunciation*. Advanced Online Publication.
- Taylor Reid, K., Trofimovich, P., & O'Brien, M. G. (2019). Social attitudes and speech ratings: effects of positive and negative bias on multiage listeners' judgments of second language speech. *Studies in Second Language Acquisition, 41*, 419–442.

- Taylor Reid, K., Trofimovich, P., & O'Brien, M. G. (in review). Using perspective taking to mitigate negative and positive bias in rating second language accent and comprehensibility. *Language and Speech*.
- Thomson, R. I. (2013). Accent reduction. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 8–11). Hoboken, NJ: Wiley-Blackwell.
- Thomson, R. I. (2018). Measurement of accentedness, intelligibility, and comprehensibility. In O. Kang & A. Ginther (Eds.), *Assessment in second language pronunciation* (pp. 11–29). New York, NY: Routledge.
- Trofimovich, P., Nagle, C. L., O'Brien, M. G., Kennedy, S., Taylor Reid, K., & Strachan, L. (2020). Second language comprehensibility as a dynamic construct. *Journal of Second Language Pronunciation*. Published online 8 July 2020.
<https://doi.org/10.1075/jslp.20003.tro>
- Tyler, A., & Bro, J. (1992). Discourse structure in nonnative English discourse: The effect of ordering and interpretive cues on perceptions of comprehensibility. *Studies in Second Language Acquisition*, *14*, 71–86.
- Weyant, J. M. (2007). Perspective taking as a means of reducing negative stereotyping of individuals who speak English as a second language. *Journal of Applied Social Psychology*, *37*, 703–716.