



## **Comprehensibility – A useful tool to explore listener understanding**

Sara Kennedy and Pavel Trofimovich

Kennedy, S., & Trofimovich, P. (2019). Comprehensibility: A useful tool to explore listener understanding. *Canadian Modern Language Review*, 75, 275–284.  
<https://doi.org/10.3138/cmlr.2019-0280>

In this review article, we discuss the construct of second language (L2) comprehensibility, which refers to listeners' judgments of ease or difficulty in understanding L2 speech. We first briefly describe intelligibility and comprehensibility, two measures which capture listener understanding of L2 speech. We then present evidence which supports the value of comprehensibility as a measure of understanding. We outline findings which show that L2 comprehensibility can be socially influenced, dynamic, and linked to multiple elements, both language-related and other. Finally, we discuss ways in which researchers, teachers, as well as speakers and listeners can approach L2 comprehensibility in ways that account for its nuanced and complex nature.

Over 30 years ago, focusing on listeners' understanding (intelligibility) of second language (L2) speech, Fayer and Krasinski (1987) categorized the study of listener understanding in two ways. The two categories were objective intelligibility, where listeners complete tasks such as transcription, showing their interpretation of L2 speakers' message, and subjective intelligibility, where listeners judge their understanding of L2 speech, typically through a scalar rating. This distinction was useful because, at the time, the term *intelligibility* encompassed both objective and subjective dimensions of listener understanding, which meant that there was no clear basis for differentiating between speech that listeners could accurately interpret or reproduce and speech that they found difficult to understand. However, within a decade, Munro and Derwing (1995a) made a convincing case for two separate yet related constructs of intelligibility and comprehensibility, which clarified Fayer and Krasinski's distinction between objective and subjective intelligibility.<sup>1</sup> In Derwing and Munro's framework, intelligibility is defined as "the extent to which a speaker's message is actually understood by a listener" (Munro & Derwing, 1995a, p. 76), and comprehensibility refers to listeners' "judgments on a rating scale of how difficult or easy an utterance is to understand" (Derwing & Munro, 1997, p. 2), two definitions we adopt here. Our goal in this review article is to highlight the value of comprehensibility as a measure in L2 speech and pronunciation research, to lay out empirical support for the complex nature of comprehensibility, and to outline future directions for working on comprehensibility in research and pedagogical contexts. Our specific objective is to demonstrate that there are sound reasons for using comprehensibility as a measure in L2 speech

---

<sup>1</sup>The research featured here builds on the foundational work by Derwing and Munro. We value these researchers' significant contributions to research and thank them for their continued mentorship of scholars. We recognize that the term comprehensibility has been used to denote the degree to which listeners understand word/utterance meaning. In this article, however, comprehensibility should be interpreted within Munro and Derwing's framework.

and pronunciation research and that researchers, teachers, raters, speakers, and listeners benefit from nuanced approaches to examining, assessing, and enhancing L2 speakers' comprehensibility.

### **1. A (brief) case for comprehensibility as a measure of listener understanding**

In their initial research, Munro and Derwing (1995a) and Derwing and Munro (1997) showed that listeners' actual understanding of L2 speech and their rated ease of understanding of the same speech were partially independent. That is, although listeners who transcribed L2 speech accurately (high intelligibility) generally also rated that speech as being easy to understand (high comprehensibility), some listeners with near-perfect transcription scores rated the speech comparatively less easy to understand.

Given this lack of direct equivalence between intelligibility measures and comprehensibility ratings, researchers and teachers of L2 speech may be tempted to avoid judging L2 speakers' comprehensibility and instead focus on measuring speakers' intelligibility. However, comprehensibility judgements can be very useful as measures of listener understanding in many contexts. One clear advantage of comprehensibility judgements is practicality. Over a relatively short period of time, listeners can judge the comprehensibility of multiple L2 speakers producing either the same or different utterances. These judgements, when made in scalar form, are quite reliable, with measures of rater consistency or agreement typically reaching high values (Munro, 2018). In contrast, measures of intelligibility require tasks which elicit listeners' actual understanding of L2 speech, unique speech content for each instance when intelligibility is measured (to avoid greater intelligibility for repeated content), and time for listeners to complete the tasks. Intelligibility scores are not highly reliable 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). Although practicality is not the only consideration, it is an important factor in designing research and assessing L2 speech.

Comprehensibility measures are not only practical and reliable, they can also provide valuable information about listeners' understanding of and reactions to L2 speech. Although comprehensibility ratings are not directly equivalent to intelligibility measures, they can show similar patterns in listeners' understanding. For example, in Kennedy and Trofimovich (2008), significant decreases in intelligibility scores across different transcription tasks were mirrored by ratings of significantly decreased comprehensibility. Similarly, in Sheppard, Elliott, and Baese-Berk (2017), higher intelligibility scores for semi-spontaneous speech predicted higher comprehensibility ratings for longer stretches of the same speech.

Comprehensibility measures can be useful for researchers and teachers in another way; research findings show links between comprehensibility ratings, listeners' effort in processing L2 speech, and listeners' attitudinal or emotional reactions. Listeners' comprehensibility ratings reflect the time they require to process L2 speech (Munro & Derwing, 1995b). Listeners have also demonstrated negative emotional reactions towards L2 speakers if listeners' processing effort is high (Dragojevic & Giles, 2016), and may even attribute lower credibility to L2 speakers with (relatively) lower comprehensibility ratings (Lev-Ari & Keysar, 2010). Although these subjective reactions may not reflect listeners' actual understanding of speech, they are important to take into account, as listeners' negative reactions can affect their perceptions of speech or speakers. As an L2 English speaker working with native English speakers in Canada notes, "As soon as I say something that they need to make an effort to understand, they just disregard me" (Vujinović, 2017, p. 27).

## **2. Comprehensibility as a multivalent construct**

Although early researchers acknowledged a multitude of factors, in addition to pronunciation, linked to comprehensibility, only recently have studies targeted the links between comprehensibility judgements and factors other than L2 speech. This work supports a view of comprehensibility as socially driven, dynamic, and varied in the linguistic and non-linguistic elements which affect it.

### *2.1. Social influences on comprehensibility*

There is decades-old evidence of social influences on listener understanding of speech. Rubin and Smith (1990) found effects for perceived speaker ethnicity on listeners' comprehension of a lecture topic (intelligibility). Listeners who were presented an Asian guise for the speaker demonstrated less understanding than listeners who were presented a Caucasian guise for the same speaker. Much less research has been conducted on social influences on comprehensibility, but studies have shown particular listener groups rating specific speakers with lower comprehensibility than the speakers' intelligibility scores would suggest. For example, Sheppard et al. (2017) showed that university faculty members who reported negative attitudes towards the English proficiency of international students gave lower comprehensibility ratings to students' L2 English speech than did faculty members who reported positive attitudes, despite both groups of faculty members being equally accurate in transcribing speech.

In a study which aimed to manipulate listeners' social bias, Taylor Reid, Trofimovich, and O'Brien (2018) had older and younger adult native English-speaking Montrealers rate the L2 English speech of Quebec francophone speakers. Some listeners were simply asked to rate the speech (baseline group), while others, prior to rating, heard comments from a confederate researcher about the L2 English skills of francophone speakers. One group of listeners heard

positive comments, and the other group heard critical comments. Compared to the baseline listeners, the positively oriented listeners generally rated speakers more favorably, including for comprehensibility. In contrast, the negatively oriented listeners' ratings depended on their age. Younger listeners rated speakers significantly more comprehensible while older listeners rated speakers significantly less comprehensible, compared to the assessments by baseline listeners. The older a negatively oriented listener was, the lower the comprehensibility ratings s/he gave.

The ratings by positively and negatively oriented listeners may have differed because negatively oriented listeners may have had their attention focused on details, such as nonnative forms in L2 speech, while positively oriented listeners may have attended to more holistic characteristics of the speech. Regardless of the explanation, this finding was striking—a positive or a negative anecdote heard before the rating influenced listeners' comprehensibility rating of an otherwise identical speech sample. The age difference in ratings may have been related to listeners' age at the time the 1977 Bill 101 was passed (a provincial law specifying French as the official language of Quebec), which lessened the influence and employability of English monolinguals. Listeners who were teenagers or adults then may have felt less willing or able to accommodate to francophone speakers of L2 English, especially when exposed before rating to negative comments about the English speech of francophones. In contrast, listeners growing up as young multilingual Montrealers in the decades following Bill 101 may have felt more open to varied language use, regardless of interlocutor proficiency, even when exposed to negative comments. That listeners' different perspectives on society and linguacultural groups can affect listener understanding of L2 speech is something that is evident for many lay people with experience of multiple languages and cultures, but is generally not accounted for in research on L2 comprehensibility.

## *2.2. Comprehensibility as a dynamic construct*

Although longitudinal development of L2 comprehensibility has been an increasingly popular focus of research, changes in the comprehensibility of L2 speech are generally measured over time periods of weeks, months, and sometimes years (e.g., Saito, Dewaele, Abe, & In'nami, 2018). However, listeners' perception of L2 speech can change over a matter of minutes (Bradlow & Bent, 2008), yet comprehensibility has rarely been framed as a dynamic, variable process which can change in real time.

Nagle, Trofimovich, and Bergeron (2019) recently explored whether comprehensibility can be construed as dynamic, examining how raters explain their assessments as they evolve in real time. Listening to samples of L2 Spanish speakers responding to personally relevant prompts, 24 Spanish-speaking raters rated 3-minute speech samples using a computer interface which allowed them to increase or decrease the comprehensibility rating by clicking a mouse as the speech unfolded. The raters then completed a stimulated recall interview, where they commented on their thoughts while watching a video capture of their rating.

Based on their rating profiles, the raters were categorized into the groups of dynamic, semi-dynamic, or non-dynamic raters. Non-dynamic raters (the majority) increased or decreased comprehensibility ratings infrequently over a speech sample. Semi-dynamic raters increased or decreased ratings at a high frequency, but the magnitude of change was small. The two dynamic raters also changed ratings at a high frequency, with a high magnitude of change which was generally in the direction of lower comprehensibility. Most raters reported that their ratings moved in the direction of greater comprehensibility either while listening to the same sample or from one sample to another. When raters' comments focused on specific dimensions of comprehensibility, nearly half of the comments about increasing comprehensibility ratings were



about discourse, such as the content or order of ideas. Discourse also featured in comments about decreasing comprehensibility ratings, but the majority of these comments were about speakers' use of vocabulary and grammar. In the end, it was unclear whether raters' profiles were related to rating strategies or to real-time processing demands.

The raters in Nagle et al.'s study, which found mixed evidence of dynamicity in comprehensibility judgments, had completed a one-way listening task, with no possibility to interact with an interlocutor. However, interactive speech, where interlocutors are reacting in real time to other interlocutors, may be a context more amenable to changes in comprehensibility judgements. Therefore, in another study exploring the dynamic nature of comprehensibility, Taylor Reid, Strachan, Trofimovich, Kennedy, and O'Brien (2018) paired L2 English speakers from different language backgrounds, completing three interactive tasks and rating each other's comprehensibility at 2.5-minute intervals for a total of seven ratings.

When interlocutors' ratings of their partners' comprehensibility were averaged, a clear pattern emerged. Comprehensibility was initially rated high, then significantly decreased at the second rating time, then gradually climbed to reach the initial rating by the seventh interval. These mutual comprehensibility ratings were aligned with interlocutors' judgments of each other's anxiety, such that interlocutors' low self-reported anxiety was associated with high ratings of partners' comprehensibility. When the comprehensibility ratings of each pair were analyzed, the difference in ratings between the two interlocutors was relatively large at the first rating time, then became aligned and was not significantly different for the remainder of rating times. The dip in comprehensibility following the initial rating may have been due to higher cognitive demands in the second interactive task, with ratings continuing to rise as interlocutors adapted to their partners' speech and to the task, as well as moving to an easier task later. A

tentative conclusion emerging from this work is that listeners' judgements of L2 comprehensibility can change in real time according to listeners' immediate experience, particularly for listeners in interactive speaking tasks.

### *2.3. Linguistic correlates of comprehensibility*

As many speakers and listeners are aware, comprehensibility ratings also reflect the linguistic content of speech. For instance, as early as 1995, Munro and Derwing reported significant correlations between listeners' comprehensibility ratings and speakers' grammatical accuracy. Based on recent research, the linguistic elements linked to L2 speakers' English comprehensibility can be roughly separated into two dimensions (e.g., Saito, Trofimovich, & Isaacs, 2017): pronunciation (individual segments, prosody, fluency) and lexicogrammar (variety and richness of vocabulary and accuracy/complexity of grammar). However, the importance of these elements for comprehensibility can depend on the linguistic background of the L2 speaker and on the speaking task (e.g., Crowther, Trofimovich, Isaacs, & Saito, 2018).

With respect to other languages, only L2 German and L2 French have been explored for linguistic correlates to comprehensibility. For L2 French, Trofimovich, Kennedy, and Blanchet (2017) found links between comprehensibility ratings and speakers' prosody and fluency, while comprehensibility ratings in Bergeron and Trofimovich (2017) were linked to pronunciation, lexicogrammar, fluency, and discourse richness. In a study targeting L2 German, O'Brien (2014) showed that comprehensibility was tied to fluency and accuracy measures for vocabulary, morphology, and pronunciation.

Although comprehensibility ratings are generally consistent across raters, the reasons for raters' decisions can vary widely. In only a few studies have raters been asked to explain their comprehensibility ratings. Isaacs, Trofimovich, and Foote (2018) had experienced instructors of

English for academic purposes in Canada and the UK repeatedly try out increasingly refined versions of a comprehensibility rating scale, using unique L2 English semi-spontaneous speech samples in each rating session. In focus group discussions, instructors weighed pronunciation and fluency as more important than vocabulary and grammar in rating comprehensibility, and recommended keeping vocabulary and grammar as separate elements in the scale. Kennedy, Blanchet, and Gu nette (2017) also had four teacher-raters explain their comprehensibility ratings of interactive speech between pairs of L2 French speakers. Teacher-raters focused on speakers' general pronunciation, their knowledge and use of vocabulary, and whether speakers understood each other. However, different teacher raters emphasized different elements for comprehensibility, ranging from pronunciation, to segments, to task type.

### **3. Putting it together: Implications for research and pedagogy**

The clear social influences on judgements of comprehensibility have ramifications for research and pedagogy. For research, the importance of attitudes towards L2 speech means that eliciting measures of raters' attitudes will help in the analysis and interpretation of comprehensibility ratings. Moreover, any 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 (nativelike speaker or less proficient speaker), what is said prior to the ratings, and how these could influence raters. For pedagogy, initiatives which involve structured opportunities for positive contact between native and L2 speakers, explaining to native-speaking listeners some differences between their language and another language and giving structured practice in transcribing L2 speech, or having native speakers take the perspective of a non-English-speaking coworker have all shown to increase native speakers' positive attitude and empathy towards L2 speakers (e.g., Derwing, Fraser, Kang, & Thomson, 2014).

Regarding dynamicity in L2 comprehensibility, speech sample length generally has been considered appropriate if raters are ready to rate by the time the sample finishes. However, if comprehensibility ratings, especially for the same speaker, can vary over time and multiple measurements, it is important for researchers to consider how the length of a speech sample or the frequency of rating might influence its comprehensibility. Another important consideration is speakers' anxiety levels, which have been linked to listeners' comprehensibility ratings. Many researchers do try to ensure that recording sessions are not stressful for speakers, but practice tasks and speaker self-reports of anxiety could help to modulate or document the possible influence of anxiety on speakers' comprehensibility. As speakers feel less anxious, their comprehensibility may improve.

Pedagogically, the implications of potential dynamicity are also crucial for the length of interactions. Because L2 comprehensibility trends upward over the course of interactions, L2 speakers who want to work on being comprehensible or to gain confidence in their comprehensibility should be encouraged to seek out opportunities for spoken interactions which are not brief by nature. These might be found in group discussions or brainstorming sessions, interviews, workshops, and community group meetings. Confidence, or lower anxiety, can also be promoted by teachers or L2 speakers through calming or self-affirming exercises prior to and during spoken interaction. The relationship between self-reported anxiety and comprehensibility is also relevant to native-speaking interlocutors who may have little experience interacting with L2 speakers. Initiatives involving structured positive contact, perspective-taking, or explicit instruction regarding L2 speakers and their speech could help lower native speakers' initial anxiety while interacting with L2 speakers and possibly also contribute to more positive attitudes.

Finally, it is encouraging to see new research in which not just pronunciation but many different dimensions are being examined for their links to L2 comprehensibility (e.g., Saito et al., 2018). For pedagogical purposes, teachers, L2 learners/speakers, and potential listeners should be aware that comprehensibility is a construct with multiple elements, not just pronunciation. Elements that may be relevant or important should be highlighted as much as possible. For instance, the type of speaking task or genre, the speaker's use of vocabulary and grammar, the listener's level of anxiety, attitude towards or experience with L2 speech are all elements that could be linked to comprehensibility. Teachers, speakers, and listeners (including native speakers) can then work on elements over which they potentially have control, such as anxiety, pronunciation, vocabulary use, attitude, or experience with L2 speech. Clearly, neither teachers, speakers, nor listeners can attend simultaneously to all elements which could be linked to comprehensibility. The importance of particular elements will vary according to the person and context, and speakers and listeners should be encouraged to enhance their awareness of these elements through consciousness-raising activities, such as guided reflection, analysis of self or others' spoken language (beyond simply pronunciation) or body awareness (to manage anxiety).

#### **4. Conclusion**

In this article, we made a case for comprehensibility as a valuable construct, arguing that social attitudes and biases can influence listeners' judgements of comprehensibility, suggesting that comprehensibility ratings have the potential for dynamic change, especially during interactive speech, and explaining that multiple elements are linked to comprehensibility ratings. Our goal was practical, namely, to suggest different approaches to addressing social attitudes, dynamic change, and multiple elements in researching, teaching, and learning L2 comprehensibility. And although we predominantly focused on comprehensibility, we acknowledge that various

measures of listeners' understanding are crucial in examining, teaching, and learning L2 speech, and advocate an awareness of both intelligibility and comprehensibility as important constructs. Last but not least, most research discussed here was carried by Canadian researchers, often targeting users of English and French (Canada's official languages) or speakers engaged in study or work in Canada. Much of this research has also been generously funded through the Government of Canada's research grants, and several publications appeared in Canadian journals. We are immensely grateful and proud to be part of a large (and constantly growing) cohort of Canada-based scholars researching L2 speech learning.

## References

- Bergeron, A., & Trofimovich, P. (2017). Linguistic dimensions of accentedness and comprehensibility: Exploring task and listener effects in second language French. *Foreign Language Annals, 50*, 547–566.
- Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*, 707–729.
- Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2018). Linguistic dimensions of L2 accentedness 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., Fraser, H., Kang, O., & Thomson, R. I. (2014). L2 accent and ethics: Issues that merit attention. In A. Mahboob & L. Barratt (Eds.), *Englishes in multilingual context* (pp. 63–80). Berlin: Springer.

- Dragojevic, M., & Giles, H. (2016). I don't like you because you're hard to understand: The role of processing fluency in the language attitudes process. *Human Communication Research*, 42, 396–420.
- Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning*, 37, 313–326.
- 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.
- 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, 459–489.
- Kennedy, S., Blanchet, J., & Gu nette, D. (2016). Teacher-raters' assessment of French lingua franca pronunciation. In T. Isaacs & Trofimovich, P. (Eds.). *Second language pronunciation assessment* (pp. 210–235). Bristol: Multilingual Matters.
- Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, 46, 1093–1096.
- 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. (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.
- Nagle, C., Trofimovich, P., & Bergeron, A. (2019). Toward a dynamic view of second language comprehensibility. *Studies in Second Language Acquisition*.
- O'Brien, M. G. (2014). L2 learners' assessments of accentedness, fluency, and comprehensibility of native and nonnative German speech. *Language Learning*, *64*, 715–748.
- 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., 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., Trofimovich, P., & Isaacs, T. (2017). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*, 439–462.
- 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.



- Taylor Reid, K., Strachan, L., Trofimovich, P., & O'Brien, M. (October, 2018). *Second language comprehensibility as a dynamic construct: Evidence from L2–L2 interaction*. Paper presented at the Second Language Research Forum, Montreal, Quebec.
- Taylor Reid, K., Trofimovich, P., & O'Brien, M. (2018). Social attitudes and speech ratings: Effects of positive and negative bias on multiage listeners' judgments of second language speech. *Studies in Second Language Acquisition*. Published online 13 November 2018.
- Trofimovich, P., Kennedy, S., & Blanchet, J. (2017). Development of L2 French oral skills in an instructed setting: A focus on speech ratings. *Canadian Journal of Applied Linguistics*, 20, 32–50.
- Vujinovic, S. (2018). *How do three immigrant women from former Yugoslavia perceive the role of accent in intelligibility and comprehensibility in the Canadian workplace?* Unpublished manuscript, York University, Canada. <http://hdl.handle.net/10315/34196>