Thir, V. (2023). Understood or not? Issues in using orthographic transcription for assessing intelligibility to international listeners of English. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 249–260). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8225289

Understood or not? Issues in using orthographic transcription for assessing intelligibility to international listeners of English

Veronika Thir
University of Passau

The most popular method for measuring intelligibility of L2 pronunciation is orthographic transcription (Munro & Derwing, 2015). This paper discusses the construct of intelligibility underlying this method, as well as its strengths and weaknesses compared to other methods of assessing intelligibility, such as shadowing. It highlights issues arising when using orthographic transcription with large, linguistically diverse samples of listeners at different proficiency levels, as is often the case in English as a lingua franca (ELF) research. This is not to say that orthographic transcription should not be used for ELF intelligibility studies, in particular quantitative ones. Instead, a case is made for carefully considering its potential biases in particular research contexts, so that appropriate measures to counteract them can be taken to ensure a satisfactory level of validity and reliability. This argument is illustrated with examples from a large-scale study ($N = 508$) involving both native and non-native listeners from over 80 different L1 backgrounds at different proficiency levels.

Keywords: intelligibility, orthographic transcription, research methods, English as an international language (EIL), English as a lingua franca (ELF)

1 Introduction

Since intelligibility is frequently named as the primary goal of L2 pronunciation teaching, its assessment has particular relevance to researchers and practitioners in L2 pronunciation pedagogy. However, since it is a perceptual phenomenon, its measurement constitutes a challenge. Various methods for assessing intelligibility have been proposed, such as comprehension questions, responses to true/false statements, listener summaries, scalar ratings, shadowing or even focused interviews (see Kang et al., 2018; Munro & Derwing, 2015). The most popular one in L2 pronunciation research is orthographic transcription (Munro & Derwing, 2015), whereby subjects listen to an audio stimulus and write down what they hear. The number of correctly transcribed words (which may be minimal pairs, key words or entire sentences, depending on the research questions) is taken to reflect the amount of understanding that has taken place.

This paper takes a closer look at orthographic transcription by examining the construct of intelligibility underlying it, as well as its validity and reliability in comparison to other widespread methods of measuring intelligibility, in particular shadowing. In doing so, it scrutinises two approaches to coding orthographic transcriptions: the 'exact word match' technique and the allowance of spelling errors. The latter seems preferable with non-native listeners, who may be penalised when using a strict exact word match. However, if spelling errors are accepted, it may sometimes be difficult to distinguish them from transcriptions reflecting unintelligibility. Such 'ambiguous' transcriptions become more frequent when working with large, linguistically heterogenous listener samples, as is the case in quantitative studies on international intelligibility, i.e., intelligibility among English as a lingua franca (ELF) users. The present paper highlights issues and challenges in this respect by reporting on the process of data coding in an ELF intelligibility study which involved over 500 native and non-native listeners from a wide variety of L1 backgrounds at different proficiency levels. It is argued that neither coding technique is inherently superior, but that careful consideration of the sample and research context is necessary before deciding on a technique, which also needs to be considered in the interpretation of results.

2 What type of intelligibility are we measuring with orthographic transcription?

Intelligibility is a complex concept that is defined in various ways within and across different fields (e.g., Derwing & Munro, 1997; Schiavetti, 1992) and different methods of assessing intelligibility might more validly reflect certain intelligibility constructs than others. When listeners transcribe the words they hear, they are essentially engaging in the process of word identification. This most closely resembles Smith's (1992) definition of *intelligibility* as "word/utterance recognition" (p. 76) in his framework; the other two dimensions are *comprehensibility* (i.e., understanding the literal meaning of a word or utterance) and *interpretability* (i.e., understanding the pragmatic meaning behind a word or utterance). One weakness of this framework, however, is its implicit bottom-up view of the comprehension process: the three components are regarded as "degrees of understanding on a continuum, with intelligibility being lowest and interpretability being highest" (Smith, 1992, p. 76). This leads to issues in empirical operationalisation, since

> the comprehension of speech is not a linear process: when a word cannot be recognized on the basis of bottom-up cognitive processes, the listener may nonetheless be able to 'fill it in' by exploiting top-down strategies. (Munro & Derwing, 2015, p. 378)

Thus, orthographic transcription "reflects more than just low-level speech processing" (Munro & Derwing, 2015, p. 378), i.e., it measures more than intelligibility as defined by Smith. Zielinski (2012) also points to "the highly interdependent nature of the different components of the Smith framework" (p. 405) and, likewise, to difficulties in empirical application:

> […] researchers might find it difficult to measure intelligibility as separate from comprehensibility because listeners will use what is a natural process of listening and refer to the meaning and context of an utterance to identify difficult words within it. (Zielinski, 2012, p. 405)

There is abundant empirical evidence supporting the use of context in the process of spoken word recognition (see e.g., Bent et al., 2019; Kennedy & Trofimovich, 2008; Thir 2020a, 2020b). Thus, if we interpret the components in Smith's framework as interrelated rather than hierarchical, this more accurately reflects the nature of speech comprehension. Intelligibility defined as spoken word recognition is therefore best understood as a cognitively interactive process that involves going back and forth between bottom-up and top-down processing (see also Mirman, 2017). Following Magnuson (2017), intelligibility is defined here as the process of "[…] map[ping] phonological forms to intended words in memory" (p. 76). With these adjustments, orthographic transcription seems well suited to measuring intelligibility as spoken word recognition (henceforth SWR). It also seems to meet quality criteria in academic research, such as validity and reliability.

3 Validity and reliability of orthographic transcription in measuring intelligibility as SWR

Validity, here mainly understood as construct validity (e.g., Gass, 2015), refers to the quality of accurately capturing the phenomenon under investigation: "a test is valid if it measures what it is supposed to measure" (Dörnyei, 2007, p. 51). One major advantage of orthographic transcription is that it allows SWR to be quantified more accurately than, for instance, listener ratings of intelligibility. Listener ratings lack validity since they only reflect how much listeners *think* they understood (Munro & Derwing, 2015) and risk confounding intelligibility with similar but separate phenomena such as "ease or effort of understanding" (termed comprehensibility in Munro & Derwing, 2015, p. 380). Compared to methods such as comprehension questions, summaries, and sentence verification, orthographic transcription allows for a more precise focus on the phenomenon of SWR. Moreover, it makes it possible to measure the intelligibility of decontextualised words (e.g., by presenting individual words or nonsense sentences), which may be necessary in certain research contexts (see §4).

One might argue that the same is true for shadowing, another common technique to measure SWR. This involves listeners instantly repeating an audio stimulus, with correct repetition being treated as signifying understanding. However, it is debatable whether correct repetition can be regarded as a sign of correct word identification (i.e., of having assigned the auditory input to the intended entry in the mental lexicon), or whether it measures phoneme (or even phone) recognition. Clearly, it is possible to correctly repeat a sound sequence without understanding what it means. At least in the case of English, it is, however, more difficult to correctly transcribe a word without knowing what the sound sequence refers to, so the danger of measuring sound rather than word recognition seems lower in the case of orthographic transcription.

For both methods, a relevant question is how precisely they measure SWR: do they only measure listeners' ability to identify spoken words, or do other skills factor into them?

Orthographic transcription obviously presupposes some competence in standard orthography, and the extent to which it also measures orthographic skills depends on how potential spelling errors are treated in data analysis. When examining previous research, two approaches emerge in this respect. The exact word match (e.g., Derwing & Munro, 1997; Munro et al., 2006) only accepts perfect matches to standard orthography. The allowance of spelling errors[1] (e.g., Bradlow & Pisoni, 1999; Field, 2005; Kennedy & Trofimovich, 2008), however, arguably offers greater validity since it does not conflate orthographic skills and SWR. This seems particularly important when working with non-native listeners and other groups prone to orthographic weaknesses, who might be penalised when a strict exact word match is used. However, the allowance of spelling errors does not necessarily meet the criterion of reliability, though this may be counteracted to some extent by using multiple coders when analysing listeners' transcriptions.

Hence, some may be quick to recommend shadowing in lieu of orthographic transcription when working with non-native listeners, to overcome the issue of additionally measuring orthographic skills. However, shadowing presupposes something else which many L2 listeners lack: an ability to articulate target words to a recognisable extent. As L2 pronunciation researchers and practitioners can well attest, numerous non-native listeners may very well correctly identify a word or phrase but may not necessarily be able to pronounce it in a recognisable way themselves. Thus, shadowing may lack validity in that it measures not only SWR, but also pronunciation skills. This is particularly problematic when studying the intelligibility of minimal pair words, which L2 listeners might be able to distinguish perceptively, but at the same time might not be able to distinguish productively. A similar issue may arise when working with native listeners whose L1 dialect contains different phonemic distinctions than the variety under study. Since orthography is far more standardised than pronunciation, and L2 learners are typically more competent in spelling than pronunciation, especially when it comes to minimal pair words, orthographic transcription seems to offer greater validity in such cases.

The second criterion, reliability, refers to consistency in measurement (see, for example, Dörnyei, 2007). Regarding raters or coders, it may be defined as "degree of consistency with which instances are assigned to the same category by different observers or by the same observer on different occasions" (Hammersley, 1992, p. 67), which refers to inter- and intra-rater reliability respectively. These types of reliability may be improved by using predetermined, precisely defined criteria when coding or rating participant responses. However, depending on the type of responses, a certain amount of personal judgment by the rater may still be necessary, potentially compromising inter- and intra-rater reliability. One major advantage of orthographic transcription is that it typically necessitates less such judgment than other measures of intelligibility that involve a greater amount of open-ended input by respondents, such as summaries or comprehension questions. Shadowing does not fare better either, since, provided no acoustic analysis is performed, it involves the auditory impressions of raters. Orthographic transcription is superior in this respect especially when using the exact word match technique; comparing participant entries to standard orthography is a straightforward and unambiguous approach ensuring high inter-rater and intra-rater reliability.

With the allowance of spelling errors, however, some variation in raters' judgments is to be expected. The extent to which inter-rater reliability may thus be compromised depends on

[1] The distinction between these approaches is not entirely clear cut. For example, Bent and Bradlow (2003) measured intelligibility by "counting the number of keywords transcribed perfectly" (p. 1605), with missing or superfluous morphemes resulting in words being coded as incorrect. However, "obvious spelling errors were not counted as incorrect" (p. 1605). Similarly, Derwing et al. (2002) describe their coding scheme as "an exact word match technique", but with the proviso that they "ignore[d] spelling errors" (p. 252).

factors such as the amount of orthographic variation in the data and the proficiency level of listeners. With moderately sized samples of educated native speaker listeners or highly proficient L2 listeners, such variation is typically comparatively minor, and inter-rater reliability is unlikely to be unacceptably low. A different picture emerges when investigating large, linguistically heterogenous samples of L2 listeners at different proficiency levels, who may all bring their own L1-specific orthographic weaknesses with them. This is particularly relevant to the ever-growing field of ELF intelligibility research, which, by definition, aims at studying intelligibility for internationally diverse populations consisting mostly of L2 listeners. In such cases, researchers may face several challenges when coding orthographic transcriptions.

4 Problems with the allowance of spelling errors: A large-scale ELF intelligibility study

Before discussing the issues that arose in the process of coding orthographic transcriptions in my own study (see §4.2), we need to consider the research context, i.e., the study's aim and underlying hypothesis, the exact method adopted, and the nature of the sample of listeners (see §4.1). This is necessary to comprehend the significance and potential implications of the issues arising in the coding process for the study's findings.

4.1 Research context

The study investigated the effect of co-text and context on the international intelligibility of two features of an Austrian accent in English (for a rationale, see Thir, 2020a, 2020b). These were the realisation of TRAP /æ/ as [e] (thus conflating the TRAP – DRESS distinction) and the realisation of NURSE /ɜː/ as [øə].[2] A subsidiary aim of the study was to investigate the effect on intelligibility of these two sound substitutions (Thir, 2020a), as well as differences in intelligibility between mono- and disyllabic words, since longer words have been found to be more intelligible than shorter words (see e.g., Howes, 1957).

A listening experiment consisting of four different conditions was developed. Three of these involved the presentation of target words in sentence co-text in the form of a cloze test (for details see Thir 2020a, 2020b), while the remaining one was a control condition where words were presented in isolation. In the conditions involving a cloze test, participants saw a single sentence on their screens with a gap in the place of the target word. In the SYN CONDITION, carrier sentences were semantically neutral and merely indicated the part of speech (POS) of the target word (e.g., *It's quite* _____.), i.e., they contained a syntactic cue. In the SYN+SEM CONDITION, in addition to indicating the POS of the target word, carrier sentences also included a semantic cue (e.g., *feather* for the target word *bird* in *They found the feather of a* _____.). In the SYN+SCH CONDITION, carrier sentences were semantically neutral but an additional schematic cue in the form of a short description under which the statement was made was presented before each sentence (e.g., *At the airport* before the carrier sentence *I need to pick up my* _____., for the target word *bag*).

Intelligibility was measured as orthographic transcription, i.e., after hearing the entire sentence, participants had to type the target word into the gap on their screens before moving on to the next sentence. Each condition contained six different types of target words which were intermixed with nine distractor words, as shown in Table 1:

[2] These are Wells' (1982, p. 120) standard lexical sets for English vowels: TRAP: /æ/, DRESS: /e/, and NURSE: /ɜː/, which are used as keywords in the main text.

Table 1

The 24 Target Words in the Experiment

CONDITION	NURSE monosyll.	NURSE disyll.	TRAP monosyll. non-MP	TRAP disyll.	TRAP MP 'different POS'	TRAP MP 'same POS'
C	*birth*	*worker*	*rat*	*palace*	*sand* (n) – *send* (v)	*land* (v) – *lend* (v)
SYN	*nurse*	*curtain*	*flat*	*massive*	*bad* (adj.) – *bed* (n)	*gas* (n) – *guess* (n)
SYN+SEM	*bird*	*purple*	*van*	*chapter*	*dad* (n) – *dead* (adj./adv.)	*pan* (n) – *pen* (n)
SYN+SCH	*firm*	*servant*	*cab*	*servant*	*bag* (n) – *beg* (v)	*pants* (n pl.) – *pence* (n pl.)

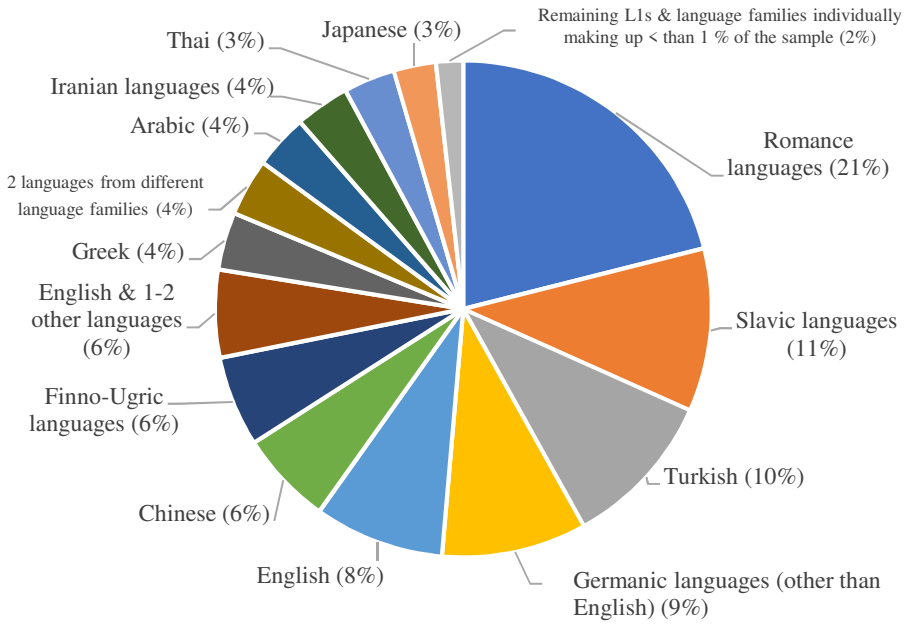
Note. For minimal pair (MP) words, the corresponding DRESS word and the part of speech (POS) is provided where n = noun, v = verb, adj. = adjective, adv. = adverb.

To obtain data from participants who were not necessarily familiar with the Austrian accent, the experiment was conducted via the internet with the help of the survey tool *SoSciSurvey*[3]. Listeners were recruited via e-mail, social media and the author's international contacts. The sample consisted of 508 (M = 175, F = 330, Other = 3) native listeners (NL: n = 66; 13%) and non-native listeners of English (NNL: n = 442; 87%), aged 18–74 years (M = 29.4). The listeners came from 81 different L1 backgrounds[4], as summarised in Figure 1. Most had a Romance language as their L1 (n = 107; 21%), followed by Slavic languages (n = 54; 11%), Turkish (n = 52; 10%) and a Germanic language other than English (n = 48; 9%).

[3] SoSciSurvey https://www.soscisurvey.de/
[4] Different combinations of two or three first languages were counted as different L1 backgrounds; for a full list, see Thir (2020b).

Figure 1

Overview of L1 Backgrounds Grouped into Selected Language Families in the Sample



Participants were asked to assess their listening proficiency using a slightly adapted version of the self-assessment scale for listening in a foreign language of the *Common European Framework of Reference* (Council of Europe, 2018, p. 167). Table 2 shows that most listeners considered themselves quite advanced (67% combined at C1 or C2 level), some reported being at an intermediate level (29% combined at B1 or B2) and the remaining participants assessed themselves at a beginner level (4% combined at A1 or A2).

Table 2

Participants' self-assessed listening proficiency

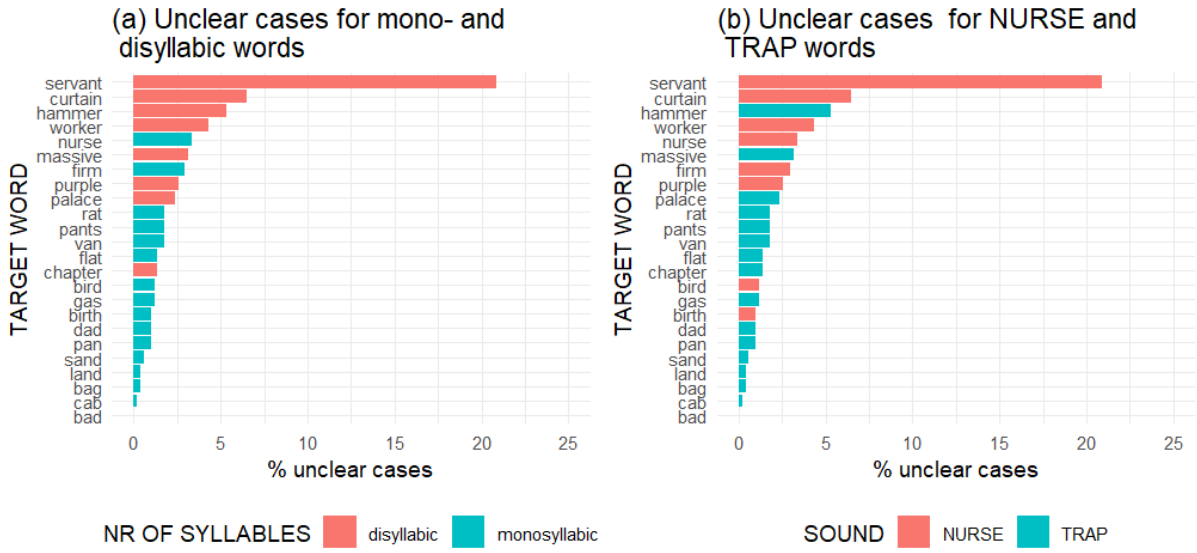
CEFR level	*n*	%
A1	8	2
A2	12	2
B1	52	10
B2	95	19
C1	129	25
C2	212	42
Total	*508*	

4.2 Issues in coding orthographic transcriptions

To increase the validity of the chosen measurement of intelligibility, the option of allowing spelling errors and missing or added morphemes was initially considered in the coding process. As a first step, 217 ambiguous transcriptions were identified, i.e., although they did not perfectly match the dictionary entry of a word, they could potentially be classified as indicative of intelligibility, which amounts to 1.8% of all 12.192 entries. Notably, certain target words exhibited a much higher percentage of such unclear cases than others. Figures 2a and 2b list all target words in descending order according to their proportion of unclear cases. The order is the same in Figure 2a (which compares mono- and disyllabic words) and Figure 2b (which compares NURSE and TRAP words). By far the highest proportion of unclear cases was identified for *servant* (20.9%), followed by *curtain* (6.5%) and *hammer* (5.3%).

Figures 2a and 2b

Percentage of Unclear Cases for Each Target Word: a) Mono- vs. Disyllabic Words; and b) NURSE vs. TRAP words



As shown in Figures 2a and 2b, there was a tendency for disyllabic words and NURSE words to exhibit a higher number of ambiguous transcriptions. Since differences in intelligibility between mono- and disyllabic words and between NURSE and TRAP words were part of the research questions investigated in Thir (2020a, 2020b), ensuring reliability in coding such cases was crucial to avoid biasing the results. Therefore, all ambiguous transcriptions were presented to eight researchers at the University of Vienna at an ELF research meeting. It soon became clear that inter-rater reliability was compromised, since there was considerable disagreement. For certain words, there was a continuum of ambiguous spellings (e.g., *purple* in Table 3), and it was impossible to decide where to draw the line between acceptable and unacceptable ones. For others, there were fewer, but equally tricky cases. For example, *van* (pronounced as [ven]), was transcribed as <vane> by six listeners. This could simply constitute a misspelling, but could also point to listeners having incorrectly identified the word as /veɪn/, transcribing it in

analogy to *pane*. Considering these circumstances, a strict exact word match approach[5] was chosen to ensure coding reliability, though this inevitably penalised (some of) the non-native listeners in the sample. Naturally, this needed to be taken into consideration when interpreting the results, especially in relation to differences in performance between native and non-native listeners, or listeners at different proficiency levels.

Table 3

Ambiguous Spellings for the Word Purple

Spelling	Frequency
purpel	1
purpal	1
purpul	1
purpule	1
purpole	1
puprle	1
perpleo	1
perpul	1
puple	1
pupple	1
purper	2
perpur	1

Note. The last two options were considered ambiguous since they resembled the German word *Purpur* (signifying *purple*).

There were, however, a few exceptions to this strict approach, which seemed clear cases of intelligibility. These included capitalisations (e.g., *Worker* or *wORKER* for *worker*), insertion of punctuation marks or of a numeral key (e.g., *flat4* for *flat*) and if parts of a carrier sentence had additionally been transcribed (e.g., *a van* for *van).* In one case, a listener had noticed the phonological ambiguity in the stimulus and provided two options (*Land* and *Lent* for *land* in the control condition), which was also accepted.

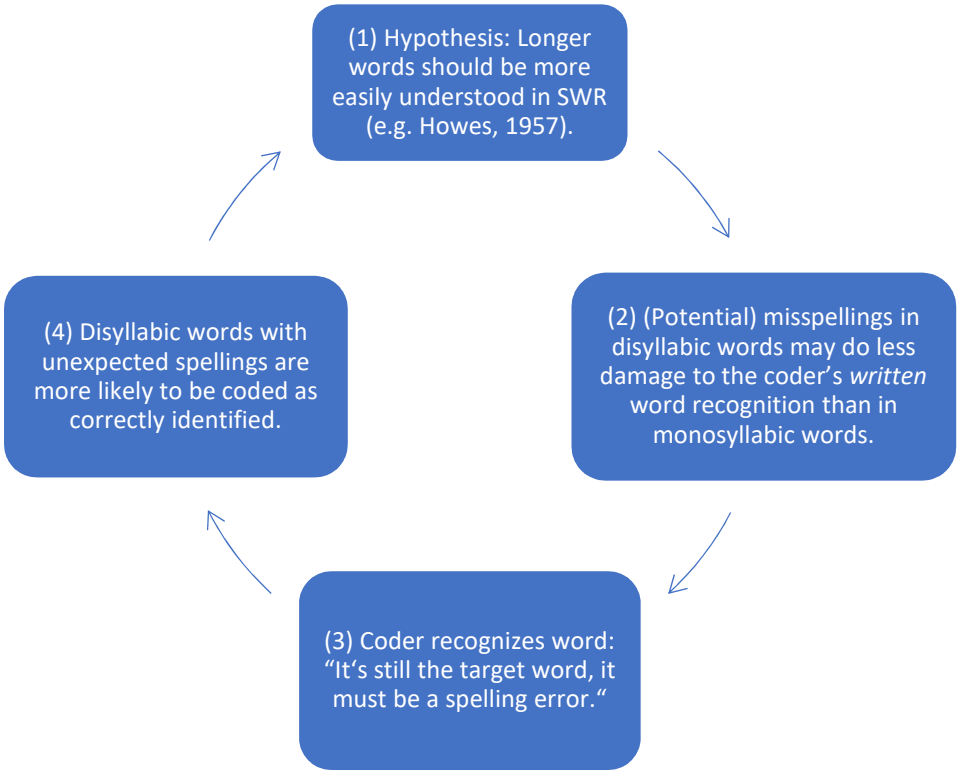
Another reason why the exact word match seemed most appropriate for this study relates to the issue of objectivity. During the coding process, it seemed more tempting to accept spelling errors in longer (i.e., disyllabic) words, since they were still more easily recognisable as the intended words. For example, it seemed obvious that **messive* was a misspelling of *massive*, while **ret* for *rat* or **lend* for *land* clearly seemed to indicate incorrect SWR, although all cases involved the substitution <e> for <a>. Especially in the case of TRAP minimal pair words, there seemed to be little reason to classify such substitutions as anything other than incorrect word identification. Since differences in intelligibility between mono- and disyllabic words and between minimal pair and non-minimal pair TRAP words were part of the research questions examined, treating the substitution <e> for <a> differently depending on the word type would have biased the results. Thus, there would have been a danger of circularity, or of falling prey to a self-fulfilling prophecy by using a coder's *written* word recognition of the target word as

[5] The choice was made in consultation with Prof. Barbara Seidlhofer, a leading scholar in the field of ELF.

a criterion in coding transcriptions. Figure 3 illustrates this issue using the example of the hypothesis that spoken longer words are more intelligible than shorter ones (caption 1). Partly influenced by this hypothesis, and partly influenced by their ability to recognise the intended word more easily in the case of disyllabic (i.e., longer) words (caption 2), the coder will likely conclude that an ambiguous transcription of a disyllabic word is a manifestation of orthographic weakness rather than failed SWR (caption 3). Consequently, disyllabic words are more likely to be classified as correctly identified than monosyllabic words (caption 4), which then results in the confirmation of the initial hypothesis (caption 1).

Figure 3

Example of Circularity in the Coding Process When Coder's Written Word Recognition Is Used as Criterion for Classifying Transcriptions



5 Discussion and conclusion

This paper highlighted several advantages of orthographic transcription to measure intelligibility if the construct is defined in terms of SWR, notably over the method of shadowing, which seems problematic when working with non-native listeners. At the same time, it stressed that the validity and reliability of orthographic transcription depends on which approach to coding is adopted: the exact word match or the allowance of spelling errors, with the former offering greater reliability but lower validity than the latter. The present paper highlighted how reliability and objectivity might be compromised when allowing spelling errors while working with large, linguistically heterogeneous samples of listeners at various proficiency levels, as is often the case in quantitative ELF intelligibility studies. Crucially, the

conclusion that should be drawn from this analysis is not that the exact word match technique is superior per se. Rather, it is necessary to carefully weigh the considerations discussed here in each particular research context, that is, in relation to the research questions and hypotheses examined, the nature of the stimuli material (e.g., minimal pair words) and the size and nature of the sample. Clearly, there will always be a trade-off between validity, reliability and also feasibility, and the exact nature of this trade-off can only be determined in each specific research context and needs to be taken into consideration when interpreting the results.

This paper also demonstrated that even in quantitative studies, some issues in data analysis might only emerge when taking a closer look at one's raw data, such as what participants typed in as transcriptions. This highlights the importance of poring intently over one's raw data, especially when working with under-researched populations for whom certain well-established research methods might turn out to be (somewhat) problematic. International listeners, especially at intermediate and lower proficiency levels, constitute such a population in quantitative intelligibility research. However, it is precisely these listeners who need to receive greater attention due to their ever-increasing use of English for cross-cultural communication worldwide and the ensuing issues of intelligibility in international contexts. This is not to say that well-established methods should not be used with such populations, but that certain adaptations might be necessary to ensure an appropriate balance of validity and reliability.

References

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America*, *114*(3), 1600–1610. https://doi.org/10.1121/1.1603234

Bent, T., Holt, R. F., Miller, K., & Libersky, E. (2019). Sentence context facilitation for children's and adults' recognition of native- and nonnative-accented speech. *Journal of Speech, Language, and Hearing Research, 62*(2), 423–433. https://doi.org/10.1044/2018_JSLHR-H-18-0273

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America, 106*(4), 2074–2085. https://doi.org/10.1121/1.427952

Council of Europe. (2018). *Common European framework of reference for languages: Learning, teaching assessment*. Council of Europe. https://rm.coe.int/cefrcompanion-volume-with-new-descriptors-2018/1680787989

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19*(1), 1–16. https://doi.org/10.1017/S0272263197001010

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, *39*(3), 399–423. https://doi.org/10.2307/3588487

Gass, S. (2015). Experimental research. In B. Paltridge & A. Phakiti (Eds.), *Research methods in applied linguistics: A practical resource* (2nd ed., pp. 101–117). Bloomsbury Academic.

Hammersley, M. (1992). *What's wrong with ethnography? Methodological explorations*. Routledge.

Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *The Journal of the Acoustical Society of America, 29*(2), 296–305. https://doi.org/10.1121/1.1908862

Kang, O., Thomson, R., & Moran, M. (2018). Empirical approaches to measuring the intelligibility of different varieties of English in predicting listener comprehension. *Language Learning, 68*(1), 115–146. https://doi.org/10.1111/lang.12270

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*(3), 459–489. https://doi.org/10.3138/cmlr.64.3.459

Magnuson, J. S. (2017). Mapping spoken words to meaning. In G. Gaskell & J. Mirkovic (Eds.), *Speech perception and spoken word recognition* (pp. 76–96). Routledge.

Mirman, D. (2017). Zones of proximal development for models of spoken word recognition. In G. Gaskell & J. Mirkovic (Eds.), *Speech perception and spoken word recognition* (pp. 97–115). Routledge.

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition, 28*(1), 111–131. https://doi.org/10.1017/S0272263106060049

Munro, M. J., & Derwing, T. M. (2015). Intelligibility in research and practice: Teaching priorities. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 377–396). Wiley-Blackwell. https://doi.org/10.1002/9781118346952.ch21

Schiavetti, N. (1992). Scaling procedures for the measurement of speech intelligibility. In R. D. Kent (Ed.), *Intelligibility in speech disorders* (pp. 11–34). John Benjamins.

Smith, L. E. (1992). Spread of English and issues of intelligibility. In B. B. Kachru (Ed.), *The other tongue: English across cultures* (2nd ed., pp. 75–90). University of Illinois Press.

Thir, V. (2020a). International intelligibility revisited: L2 realizations of NURSE and TRAP and functional load. *Journal of Second Language Pronunciation, 6*(3), 458–482. https://doi.org/10.1075/jslp.20012.thi

Thir, V. (2020b). *Phonological intelligibility in ELF communication: The role of co-text and context* [Unpublished doctoral dissertation]. University of Vienna.

Wells, J. C. (1982). *Accents of English*. Cambridge University Press.

Zielinski, B. (2012). Intelligibility in World Englishes: Theory and Application by C. L. Nelson. *World Englishes, 31*(3), 404–406. https://doi.org/10.1111/j.1467-971X.2012.01765.x

About the author

Veronika Thir holds a PhD in English Applied Linguistics from the University of Vienna. Her research interests focus on issues of intelligibility in ELF communication, in particular as they relate to the use of linguistic co-text and extralinguistic context by international listeners. Her work has been published in the *Journal of Second Language Pronunciation* and as a chapter in *Language Change: The Impact of English as a Lingua Franca* (2021, CUP). She is currently working as a post-doctoral researcher at the University of Passau, in a research project on cultural differences regarding personal data disclosure.

Email: Veronika.Thir@uni-passau.de