Social Sciences & Humanities Open 6 (2022) 100313

Contents lists available at ScienceDirect

Social Sciences & Humanities Open

journal homepage: www.sciencedirect.com/journal/social-sciences-and-humanities-open

Regular Article

Testing the Lingua Franca Core: The intelligibility of flaps

Darío Barrera-Pardo

Universidad de Sevilla, Departamento de Lengua Inglesa, c/ Palos de la Frontera, s/n, 41004, Sevilla, Spain

ARTICLE INFO

Keywords:
English as a lingua franca
Lingua franca core
Pronunciation teaching
General American
Flaps
Word frequency
Experience with L2

ABSTRACT

Over twenty years ago, Jenkins (2000) put forth the Lingua Franca Core (LFC), a pronunciation syllabus for international intelligibility among non-native speakers (NNSs) of English. Although insufficient empirical research has been directed to validating the LFC proposals, the few studies that have tested this syllabus have produced mixed findings. One of the core features of the LFC is the use of British-based /t/, rather than General American (GA) flap [ɾ], which allegedly has a negative impact on English as lingua franca (ELF) intelligibility. There are, however, three additional types of flap in accents such as GA. In the current study, the intelligibility of the four types of flap typical of GA were tested experimentally, in the context of learners at an intermediate level with four European language backgrounds ($n = 78$). Using a matched-guise technique, learners were presented with the flapped and non-flapped versions of words including the four types of flap. The results of two experiments indicate that three of the four types of flap proved detrimental for ELF intelligibility to a large extent. More expectedly, word frequency and experience with GA were moderately associated with flap intelligibility.

1. Introduction

As is well known, nonnative speakers (NNSs) of English greatly outnumber native speakers (NSs) (e.g. Graddol, 2006; Jenkins, 2000, 2009). In this context, oral dimensions important for communication, such as speech intelligibility (as defined by Munro & Derwing, 1995), should be of interest to English pronunciation teachers and researchers worldwide. The lingua franca core (LFC) proposed by Jenkins (2000, 2002, 2007), a set of pronunciation features that allegedly preserve mutual intelligibility among NNSs, has generally (but far from unanimously, see e.g., Dauer, 2005; Szpyra-Kozłowska, 2008) met with enthusiastic reactions from teachers (Levis, 2016; Szpyra-Kozłowska, 2015; Tanner & Henrichsen, 2022). Besides the enthusiastic reaction to the LFC manifested by many pronunciation practitioners and in several circles (such as teachers' associations), the significance of this line of research is well-attested by the impact that Jenkins' (2002) paper has had on the pronunciation literature. Thus, Demir and Kartal's (2022) bibliometric analysis showed that Jenkins (2002) is one of the most cited papers in the L2 pronunciation research literature.

One of the strengths of the LFC proposal lies in its answer to the fact that interactions among NNSs can be characterized by their fluid and dynamic phonology. Rather than constituting speech communities, these lingua franca speakers can be thought of as integrating what have been called 'Transient International Groups' (TIGs) (see Pitzl, 2018 for more details). These TIGs are mainly defined by not having a shared repertoire, that is, a common phonology of some kind, but by showing ephemeral interactions with no easily observed common phonological features (O'Neal, 2020). While a discussion of this important matter is beyond the scope of the current study, it should be noted that insufficient attention to individual variation has been paid in current L2 pronunciation research (Munro & Derwing, 2015; Munro, Derwing, & Holt, 2012). Learners sharing the same L1 may diverge markedly in their phonological learning trajectories, and the ephemeral interactions that have been attributed to lingua franca communication are in all likelihood shaped to some degree by the phenomenon of individual variation. The significance of the LFC is precisely that it offers a starting point for describing which intelligible phonetic features emerge from the interactions among NNSs; there are, however, other conceptions of EFL research, such as O'Neal's (2019) view that "pronunciation in ELF interactions is better described as variation and flexibility rather than as a set of core features" (p. 121), and one potential cause for this variation is the individual variability so often overlooked in L2 pronunciation. Set against the backdrop of the LFC as an answer to the idiosyncrasy of lingua franca communication, it seems urgent to substantiate or refute the proposals advanced in this pronunciation syllabus. To reiterate and as O'Neal (2020) affirms, "to ignore the phonology of such interactions is to ignore the most common usage of English" (p. 239). Empirical approaches to the LFC are hence called for.

Jenkins' LFC places emphasis on segmentals rather than on supra-segmental aspects of pronunciation; among the consonant targets that

E-mail address: dario@us.es.

https://doi.org/10.1016/j.ssaho.2022.100313
Received 22 February 2022; Received in revised form 29 May 2022; Accepted 23 June 2022
Available online 8 July 2022

favor intelligibility among NNSs is British English /t/ instead of the American English flap (which is transcribed [ɾ]), as in the word 'city.' In the standard accent General American (GA), these flaps are pervasive (Avery & Ehrlich, 1992; Picard, 2007), and in point of fact, as shown below, there is more than one kind of flap in GA and other accents. Given the widespread occurrence of this consonant variant in some English accents, and the relative importance placed on its recognition by ESL pronunciation methodologists (e.g., Avery & Ehrlich, 1992; Celce--Murcia, Brinton, Goodwin, & Giner, 2010; Lane, 2010; Picard, 2010), the flap deserves to be a focus of attention in pronunciation research and teaching. The aim of the current study is to test empirically the intelligibility of the flap in a European English as a lingua franca (ELF) context.

1.1. Literature review

Empirical research on the LFC proposals has not been extensively carried out (e.g., Munro & Derwing, 2015; Szpyra-Kozłowska, 2015), despite commentary claiming the contrary (e.g., Kiczkowiak, 2020). Nonetheless, the LFC as conceptualized by Jenkins (2000), requires what Walker (2010) has referred to as "fine-tuning the LFC" (p. 43), that is, supplementary data to confirm or disprove its claims, and subsequent adjustment of the LFC according to the findings of empirical research. Thus, it could be recommended that the broad enthusiasm that the LFC sparked in the pronunciation community should cautiously be tempered. In addition, Jenkin's LFC was originally based on a set of classroom observations, and these "are valid ways of documenting aspects of pronunciation, but they do not in themselves constitute sufficient evidence about learners' abilities or about the effects of instruction on learners' language output" (Derwing & Munro, 2005, p. 381). The criticism that the LFC is not based on a substantiated research program is often mentioned (e.g., Dauer, 2005; McCrocklin, 2012; Thir, 2020; Trudgill, 2008). There is however a small but growing number of studies that have tested the LFC syllabus, and to these we now turn.

Osimk (2009) found that aspiration of voiceless plosives (one of the requisites of the LFC) facilitates intelligibility among NNSs, and that non-canonical realizations of interdental fricatives did not interfere with intelligibility, as predicted by the LFC. The role of rhoticity, which is to be preferred in an ELF context according to Jenkins (2000, 2002), is less conclusive and demands more investigation. The L1s of the participants were all of European backgrounds, however. Kennedy's (2012) findings partially supported some of the claims of the LFC, since she reported very few examples of suprasegmental aspects that contributed to unintelligibility (2 out of 54 occurrences); in addition, consonant modifications and consonant cluster violations were observed to impede intelligibility, a result that is in line with the LFC syllabus. On the other hand, Kennedy found some examples (11 out of 54) that, conflicting with the LFC, involved vowel quality as a source of unintelligibility. The participants' L1 backgrounds were more diverse than those of Osimk's study. In an analysis of ELF intelligibility in the context of nine speaker-listeners from eight Asian countries and one African country, Deterding (2013) observed that 86 per cent of misunderstandings in interactions among these participants involved pronunciation (other communication breakdowns were related primarily to grammar and lexis). Deterding found that many of these pronunciation misinterpretations involved a combination of various features (for example, an unexpected vowel quality plus the deletion of a consonant), which adds an element of uncertainty in substantiating or disproving the LFC tenets. Nevertheless, some worthwhile findings that emerge from Deterding's study merit attention. These are as follows. First, in accord with the LFC, "the greatest impact on intelligibility comes from consonants" (p. 91). Second, also in agreement with the LFC, reduction of consonant clusters caused intelligibility problems. Third, and differing with the LFC, lack of aspiration of voiceless plosives was not problematic for intelligibility. Finally, vowel quality misunderstandings were few in number, which aligns with the LFC syllabus, but the absence of vowel length distinctions did not result in unintelligibility either, which contradicts Jenkins' claims. Interestingly and more in connection with the aim of the current study, Deterding found in his data only one case where flapping was involved in unintelligibility.

Luchini & Kennedy (2013) examined sources of unintelligibility in the interactions of two L1 Hindi speakers with one L1 Spanish speaker, and found that, matching the LFC features, individual consonant deviations were obstacles to intelligibility, and aspiration of voiceless plosives was also in line with the LFC proposals, but word stress (a non-core feature in Jenkins' syllabus) was as well a source of unintelligibility. In a study of the international intelligibility of Chinese-accented English, Zhang's (2013) findings corroborated some of the LFC proposals and challenged others. Vowel quality resulted in intelligibility failures (contrary to the LFC), as did vowel length, which ratifies the LFC. Additionally, aspiration of plosives and consonant substitutions as well resulted in unintelligibility, as predicted by Jenkins.

Also in regard to vowel quality, a few studies researched the LFC claim that only the /ɜː/ vowel quality is a source of intelligibility breakdown in lingua franca contexts. O'Neal (2015) found in a small-scale study that several vowel qualities were crucial in maintaining mutual intelligibility among EFL speakers. Recently, Thir (2020) tested in laboratory conditions the intelligibility of two vowel qualities, /ɜː/ and /æ/ (the latter being a vowel quality excluded from the LFC). An L1 German Austrian speaker of English recorded words with these two vowels in four conditions, first in isolation (e.g., *birth*), followed by three conditions that provided increasingly more communicative context. Thir incorporated into her analysis the notion of functional load (FL) (Catford, 1987). FL, in its narrow sense, which is prevalent in the literature (see Sewell, 2017), measures how many segmental contrasts (e.g., minimal pairs) a consonant or vowel can sustain. It is widely recognized, for instance, that the vowels /iː/ and /ɪ/ contrast in a large number of minimal pairs, while the vowels /uː/ and /ʊ/ do not. Thir asked whether substituting the vowels /ɜː/ and /æ/ with vowels that had high FL (i.e., that participated in many minimal pairs), would result in more problematic international intelligibility. She recruited 58 NS listeners and 434 NNS listeners who came from 40 different L1s. Thir's findings ran counter to the LFC stipulations, with substitutions of /ɜː/ resulting more intelligible than replacements of /æ/, but this observation was in line with FL principles. She also found, however, that contextual support tended to counterbalance these differences. Research by Jurado-Bravo (2018) further disputed the non-core status of all vowel qualities except /ɜː/ in ELF intelligibility. She employed a listening task methodology, in which an L1 Spanish speaker of English recorded sentences that featured a vowel length or quality discrimination trial; there were 125 listeners from over 20 countries around the world, representing several L1s. The length of long vowels was instrumentally manipulated to make this phonetic feature more salient. Data analyses revealed that vowel length increased intelligibility, thus confirming the LFC; on the other hand, it was found that vowel quality results were mixed and should be taken as inconclusive.

Rahimi and Ruzrokh (2016) is the only intelligibility study were the LFC was explicitly taught and compared to a traditional NS-based syllabus. Set in an Iranian context, a control group was taught with a British accent syllabus and an experimental group was instructed with a LFC program. Production and recognition posttest results revealed that the experimental group's intelligibility was superior to the control group, with a large effect size in recognition. Lewis and Deterding (2018), for their part, examined the role of word stress in the interactions of 41 speakers from various countries in South-East Asia, observing that this non-core pronunciation feature caused some misunderstandings (e.g., *cirCLED* heard as *called*), although they acknowledged that some misplaced stresses (e.g., *calenDAR*, *inJURED*) were not unintelligible, a finding that aligns with the non-core status of word stress in the LFC syllabus. Finally, Zoghbor (2018) tested the intelligibility of 50 L1 Arabic learners of English as assessed by 18 non-Arab speakers, resulting in support of most of the LFC tenets, with the exception of the quality of

the vowel /ɜː/, rhoticity, and word stress.

Studies conducted more recently yielded results that seem to attest, on the whole, to the multifarious nature of the LFC. Gadiner (2019) examined vowel length and quality in interactions between Brunei English speakers and non-Brunei English speakers. She found that vowel productions amounted to 24% of the misunderstandings among interlocutors. Vowel length caused unintelligibility in 11% of the recorded misunderstandings, and only about 6% of the miscommunication observations involved vowel quality; these findings are overall in support of the LFC. It should be noted, however, that a number of observed misunderstandings implicate more than one phonetic feature, and thus determining which pronunciation element causes unintelligibility is not always straightforward. The same remark was made previously in the current study about the observations put forth by Deterding (2013). In this connection, it could be suggested that LFC investigations of an experimental (instead of a purely observational) nature are clearly called for; experimental manipulation may guard, to some degree at least, against these confounding analyses. An apropos study in this regard was carried out by Jeong, Thorén, and Othman (2020), who looked into the intelligibility between Malaysian English and Swedish English. Manipulating three phonetic characteristics, word stress, consonant clusters and vowel length, they found that word stress, a non-core feature of the LFC, had an impact on lingua franca intelligibility. In line with the LFC stipulations, when the Malaysian speaker altered consonant clusters and vowel length, speech became more intelligible for the Swedish listeners. Barrass, Baffoe-Djan, Rose, and Boggs (2020) examined the intelligibility and comprehensibility of Korean English when listened to by Mandarin learners of English. These researchers reported that some of their findings were in line with the LFC tenets. Thus, avoidance of epenthesis in marked consonant clusters appeared to be important phonetic factors for intelligibility, whereas vowel quality realizations were not (interestingly, with the exception of /ɜː/, a core feature for Jenkins, 2000, 2002). In fact, the researchers reported that in the lingua franca setting they investigated, turning the diphthong /əʊ/ into monophthong /o/ resulted in significantly improved intelligibility. For consonants, Barras and his colleagues found that the dental fricatives /ð/ and /θ/, substituted by /d/ and /s/ respectively, were unproblematic for intelligibility. Likewise, the replacement of /r/ by lateral-like consonants (i.e., /l/) posed no difficulty in this setting. Such finding runs counter to LFC stipulations, which hold that /r/ should be realized as /ɻ/. This last observation, nonetheless, needs to be understood in the context of the L2-L2 interactions of the study, since both Korean and Mandarin realize /r/ similarly (that is, there is a potential cross-linguistic benefit). Jeong and Thorén (2018) also investigated the replacement of /θ/ by /s/ by German speakers of English, as perceived by Swedish speakers of English. Again, their results were in accordance to the LFC, since listeners had little difficulty in transcribing the utterances that contained the consonant substitution. Finally, Suntornsawet (2019) assessed the international intelligibility of Thai-accented English by L1 Arabic, Spanish, Japanese, Portuguese, and Chinese listeners (these five L1s are the most commonly used languages in the world, according to UNESCO listings). She reported that cluster simplification, consonant devoicing, lack of final released consonant, and fully stressed unstressed vowels posed the highest thereat for intelligibility in the setting investigated.

The studies reviewed seem to suggest collectively that some of the proposals of the LFC syllabus are tenable, others are not substantiated, and finally that some of the findings are inconclusive, patently demanding further research. In other words, the results of current research on the LFC are mixed; as Pickering and Huang (2022) comment, data on ELF intelligibility remain as yet inconclusive. It is in this background that the current study aims at advancing knowledge and informing practice, investigating the role of flaps in ELF intelligibility. Since flaps are ubiquitous in some standard varieties, as will be shown, the present study is a significant and original contribution to the existing literature on the LFC.

1.2. Four types of flap

Jenkins' (2000; 2002) LFC proposal made reference to the flapped sound normally heard in the speech of GA speakers. Jenkins asserted that intervocalic /t/ as heard in the speech of British English speakers is necessary for intelligibility in ELF contexts. In GA flapping is a well-established feature; Byrd (1993) found in a corpus study of GA that 20% of all oral stops were flaps. In other standard accents such as Australian English and Irish English flapping is also common in everyday speech (Shockey, 2003).

Jenkins' LFC focused on flapping of /t/ in intervocalic position, as in 'city,' but in GA and also in Australian English there are other two phonological positions where a /t/-based flap may occur. Words like 'twenty' and 'center' are regularly pronounced with a nasal flap in accents like GA. Nasal flaps appear frequently in corpus studies of GA like that of Byrd (1993). Another position where a flap may arise is after the sonorant consonants [ɹ] and [l] (Wells, 1982). Thus, words like 'party' or 'filter' may be pronounced with a flap as well. Finally, not only /t/ may be flapped; the voiced counterpart of /t/, i.e., /d/, is also regularly flapped in intervocalic position in GA. This is also a regular feature of Australian English (Cox & Fletcher, 2017). Words like 'rider' or 'lady' are regularly pronounced in GA with a flap. To summarize, then, flaps may be of four types:

- /t/ in intervocalic position, as in 'writer'; this flap is the focus of the LFC (referred to henceforward as 'flapped /t/')
- nasal flap, as in 'center' (referred to henceforward as 'nasal flap')
- /t/ after an approximant (/ɹ/ or /l/), as in 'party' or 'filter' (referred to henceforward as 'flapped /ɹt/')[1]
- /d/ in intervocalic position, as in 'rider' (referred to henceforward as 'flapped /d/')

1.3. Word frequency

In L2 studies word frequency has often been found to be correlated with speech processing (e.g., Rau, Chang, & Tarone, 2009; Trofimovich, Collins, Cardoso, White, & Horst, 2012). It has also been determined that token frequency is closely related to flap distribution, with high frequency words resulting in more flapping than low frequency words (Patterson & Connine, 2001). Pitt, Dilley, and Tat (2011) found that these frequency effects also have an impact on flap recognition.

1.4. Experience with GA

Experience with the L2 has been extensively investigated in the realm of L2 speech learning, and it has been established that, expectedly, more L2 use is correlated with less foreign accent (Major, 2008). This relationship applies to both perception and production of non-native sounds, at the segmental and suprasegmental levels (Moyer, 2009). It seems reasonable to conclude, then, that more familiarity with and use of GA will lead to higher flap intelligibility since, as commented, flaps are ubiquitous in this accent, and therefore should be widely available to learners in terms of input (Flege, 2009).

1.5. Research questions

The LFC pronunciation syllabus proposed by Jenkins (2000, 2002) included as a core item nonflapped /t/ instead of the GA-based flapped [ɾ]. Although the LFC has been the focus of some empirical research, to date few experimental studies have either substantiated or disproved Jenkins' proposals. There is clearly a need to fill this research gap. On the other hand, there is no evidence, aside from Jenkins' own

[1] Flaps in the sequence /lt/ as in 'filter' were not included in the study due to their low occurrence.

observations on NNSs interactions, that the flap has a negative impact on ELF intelligibility. This is all the more relevant because, first, as previously stated, there are in fact four types of flap in GA (and other standard accents) that can be potential targets of unintelligibility in EFL contexts, and, second, because flaps are so abundant in some standard accents.

Thus, flap intelligibility is an area of research that patently merits attention; this leads to the first research question (RQ1) of the study:

(RQ1): To what extent do the four types of flap impact on intelligibility in an ELF context?

Since word frequency mediates lexical recognition, and, as shown, in the case of flapping this relationship has also been substantiated, this study will therefore attempt to answer the following additional research question (RQ2):

(RQ2): To what extent is word frequency related to flap intelligibility?

Finally, given that access to L2 speech input seems to correlate with both perceptual and productive accuracy, the current study will also seek to answer the following research question (RQ3):

(RQ3): To what extent is experience with GA related to flap intelligibility?

2. Experiment 1

2.1. Method

2.1.1. Participants

This study was conducted at a large public university in Spain which has a representative body of international students. Participants were recruited in this setting through personal contacts and circulating flyers to gather NNSs with a B1 (according to the CEFR) English proficiency level, i.e., an intermediate level. This resulted in a total of 48 participants with different European language backgrounds (L1 Spanish $n = 23$; L1 French $n = 6$; L1; L1 Italian $n = 12$; L1 Polish $n = 7$). Their mean age was 21 years, and there were 36 female and 12 males.

2.1.2. Material

The target words for three of the types of flap discussed in section 1.2. (flapped /t/, nasal flap, and flapped /ɹt/) are presented in Appendix A. All words (with the exception of *interview*) were disyllabic trochees (words with a stressed syllable followed by an unstressed syllable), the preferred phonological pattern for flaps (Eddington & Elzinga, 2008).

2.1.3. Procedure

A matched-guise technique (Lambert, Hodgson, Gardner, & Fillenbaum, 1960) was used in this study. A 22-year old male native speaker of French with extensive experience with GA recorded the 36 words in Appendix A in carrier sentences that provided a minimal context (e.g., 'She's going to *enter* the room'); these sentences were recorded with the flapped target word and with its nonflapped version. Sentences with distractors were also recorded (e.g., 'That picture is much *darker*') by the same speaker. This resulted in a total of 72 sentences (12 words x 3 types of flap x 2 conditions) and 16 sentences with distractors. These sentences were normalized for peak amplitude and saved as audio files. 48 different blocks of sentences with the flapped condition and nonflapped condition plus the distractors were created and randomized. Each block included 3 target words, with either the flapped or nonflapped version of the word. For example, the first block featured the words *interview* (flapped), *party* (nonflapped), and *thirty* (nonflapped), plus the 16 distractor sentences. In this way it was very unlikely that the listeners would be cued to the objective of the intelligibility task. Each sentence was recorded twice, and was orally numbered in the recording. Once all the participants were recruited, the listening task was administered. This was done in a computer room, where the participants listened through headphones to one of the 48 blocks. Before this listening task, participants practiced with 3 sentences not included in the task, to familiarize themselves with the experiment's procedure. The participants listened to the blocks at their own pace, and were instructed to fill in the blank space corresponding to the prompted word, in an orthographic transcription, on a sheet. Once they finished the task, they completed a background questionnaire.

2.1.4. Analysis

Intelligibility was measured as the percentage of correctly transcribed words in the listening task. The small number of observations in Experiment 1 (12 observations for each type of flap) suggested a non-normal distribution of the data, which was confirmed by visual inspection of boxplots and Q-Q plots (as recommended by Larson-Hall, 2015). The analyses below therefore are run in the nonparametric counterparts of parametric tests (Egbert & LaFlair, 2018). Acknowledging the recommendations of quantitative research methodologists (e.g., Cummings, 2012; especially Larson-Hall, 2015; Norris, Plonsky, Ross, & Schoonen, 2015; Plonsky, 2015, 2021), the analyses of the results reported focus primarily not on statistical significance, but on the effect sizes observed, their confidence intervals (CIs), and their practical significance. The effect sizes for the Wilcoxon signed-rank tests and Mann-Whitney U tests are reported as biserial correlations (r_b), whose coefficients according to Brown (1988) are interpreted the same as those for Pearson's *r*. Plonsky and Oswald (2014) offered the following guidelines for correlation coefficients in L2 research: "we suggest that *r*s close to 0.25 be considered small, 0.40 medium, and 0.60 large" (p. 889). For nonparametric Spearman's rho correlations, Cohen (1988) suggested these values: for a large correlation rho is 0.5, for a medium rho is 0.3 and for a small correlation rho is 0.1. The statistical analyses were run in JASP (https://jasp-stats.org/) (see e.g., Loerts, Lowie, & Seton, 2020). Learners' self-reported experience with GA was collected through three items on the background questionnaire, which had three 7-point Likert scales with the following questions: *How familiar are you with American English?* (1 = not at all familiar, 7 = very familiar); *Do you try to use an American accent?* (1 = never, 7 = always); *Do you normally listen to American English? (e.g., songs, movies, series)* (1 = never, 7 = always). Experience with GA was computed as the average of these three items.

2.2. Results

Table 1 gives descriptive statistics for the intelligibility of the three types of flap investigated in Experiment 1. As the results indicate, listeners found all three nonflapped versions easier than their flapped counterparts. The largest differences were detected for the nasal flap, followed by flapped /ɹt/, and finally flapped /t/.

Wilcoxon's signed-rank tests were run on the data. Nonflapped /nt/ ($Mdn = 100$) was significantly more intelligible than the nasal flap ($Mdn = 37.5$), $W = 76.5$, $p = 0.004$, with a very large effect size, $r_b = 0.96$, 95% CI [0.87, 0.99]. The confidence interval for this effect size is narrow and precise, which means that with 95% confidence the true effect in the population could be at least as large as 0.87, or very large, 0.99. This means that nonflapped /nt/ is much more intelligible than the nasal flap. Nonflapped /ɹt/ ($Mdn = 79.41$) was significantly more intelligible than flapped /ɹt/ ($Mdn = 31.25$), $W = 53.5$, $p = 0.009$, with a very large effect size, $r_b = 0.94$, 95% CI [0.79, 0.98]. The confidence interval for this effect size is narrow and precise, which means that with 95% confidence the true effect in the population could be at least as large as 0.79, or very large, 0.98. Nonflapped /ɹt/ is clearly much more intelligible than its flapped counterpart. Finally, nonflapped /t/ ($Mdn = 87.5$) was not significantly more intelligible than flapped /t/ ($Mdn = 75$), $W = 28.5$, $p = 0.157$, with a large effect size, $r_b = 0.58$, 95% CI [−0.11, 0.89]. The confidence interval includes zero, which means

Table 1
Descriptive statistics for Experiment 1.

	nasal	nonflapped	flapped	nonflapped	flapped	nonflapped
	flap	/nt/	/ɹt/	/ɹt/	/t/	/t/
n	12	12	12	12	12	12
Median	37.5	100	31.25	79.41	75	87.5
IQR	28.12	15.62	65.62	40.62	40.62	37.5
Minimum	0	12.5	0	50	37.5	50
Maximum	87.5	100	100	100	100	100

Note: IQR = Interquantile Range.

that the effect size is not statistically significant. In addition, the confidence interval for this effect size is very wide and therefore imprecise, with a possible difference in the population, with 95% confidence, as small as −0.11 or as large as 0.89.

A nonparametric Kruskal-Wallis test of differences among the three types of flap was conducted, resulting in x^2 (2) = 11.6, $p = 0.003$, with a medium effect size $\varepsilon^2 = 0.33$. This means that 33% of the variance in intelligibility can be attributed to the type of flap. Post-hoc comparisons revealed that the difference in intelligibility between the nasal flap and the flapped /ɹt/ was not statistically significant, $W = -0.74$, $p = 0.858$, $r_b = -0.12$, 95% CI [−0.54, 0.33]. The effect size is negligible and statistically nonsignificant (the CI includes zero). The difference in intelligibility between the nasal flap and flapped /t/ was statistically significant, $W = 3.99$, $p = 0.013$, $r_b = -0.67$, 95% CI [−0.85, −0.33]. The effect size is large, and it could be a small effect size ($r_b = -0.33$) or a large effect size ($r_b = -0.85$) in the population, with 95% confidence. The wide CI is probably due to the small sample size, indicating that it is a rather poor estimate of the true effect that might exist in the population. Finally, the difference in intelligibility between the flapped /ɹt/ and flapped /t/ was statistically significant, $W = 123$, $p = 0.003$, $r_b = 0.71$, 95% CI [0.40, 0.87]. The effect size is large, and it could be a medium effect size $r_b = 0.40$ or a very large effect size $r_b = 0.87$ in the population, with 95% confidence.

Word frequency was calculated by dividing the total sum of frequency scores in reference to the Corpus of Contemporary American English (Davies, 2019), and $\log_{10}$ transforming the scores according to Schleef (2013). Spearman's rho correlations revealed that there was a significant ($p = 0.009$) and medium correlation between word frequency and intelligibility for all the flaps, rho = 0.43. The correlation between the intelligibility of the nasal flap and word frequency was small and nonsignificant, Spearman's rho = 0.18, $p = 0.58$. The correlation between flapped /ɹt/ intelligibility and word frequency was large and statistically significant, Spearman's rho = 0.71, $p = 0.009$. Finally, the correlation between flapped /t/ intelligibility and word frequency was statistically significant and large, Spearman's rho = 0.72, $p = 0.008$. Lastly, experience with GA, as self-reported by the participants, was not related statistically to flap intelligibility, $p = 0.21$, although the correlation effect size is small, Spearman's rho = 0.13.

3. Experiment 2

To further explore the intelligibility of flapped /t/ and with a view to additionally investigate flapped /d/, Experiment 2 was designed to increase the sample size (the number of flapped and nonflapped tokens) to 30 observations.

3.1. Method

3.1.1. Participants

A different sample of participants with the same characteristics of those in Experiment 1 were recruited, resulting in 30 participants (L1 Spanish $n = 16$, L1 French $n = 3$, L1 Italian $n = 8$, L1 Polish $n = 3$). Their mean age 20.5 years, and there were 21 females and 9 males.

3.1.2. Material

The target words are presented in Appendix B.

3.1.3. Procedure

The procedures are the same as in Experiment 1, this time resulting in 30 blocks, each with one flapped /t/, one nonflapped /t/, one flapped /d/, and one nonflapped /d/, plus 18 distractor sentences.

3.1.4. Analysis

As in Experiment 1, intelligibility was measured as the percentage of correctly transcribed words in the listening task. Likewise, as the data were not normally distributed, the analyses were nonparametric.

3.2. Results

Table 2 gives descriptive statistics for the intelligibility of the two types of flap investigated in Experiment 2. There was a comparatively large difference between flapped /t/ and nonflapped /t/, and a less pronounced contrast between flapped /d/ and nonflapped /d/.

Wilcoxon's signed-rank tests resulted in the following. Nonflapped /t/ (*Mdn* = 100) was significantly more intelligible than flapped /t/ (*Mdn* = 77.77), $W = 185$, $p = 0.003$, with a large effect size, $r_b = 0.76$, 95% CI [0.46, 0.90]. The confidence interval for this effect size is relatively precise, with a possible medium effect size (0.46) or a very large effect size (0.90) in the population, with 95% confidence. Thus, the difference in intelligibility expressed by the CI of the effect size is real, ranging from medium to large. Nonflapped /d/ (*Mdn* = 88.88) was significantly more intelligible than flapped /d/ (*Mdn* = 83.32), $W = 113$, $p = 0.029$, with a large effect size, $r_b = 0.61$, 95% CI [0.16, 0.84]. The confidence interval for this effect size is very wide, therefore not much confidence can be placed in this result. The true effect size in the population, with 95% confidence, could be as small as a trivial 0.16, or very large, 0.84. A Mann-Whitney *U* test showed that flapped /t/ was not significantly more intelligible than flapped /d/, $U = 471$, $p = 0.755$, with a negligible effect size $r_b = 0.04$, 95% CI [−0.24, 0.32]. The confidence interval for this effect size is wide and not statistically significant, since it includes zero. Word frequency was computed as in Experiment 1. Overall, there was a statistically significant and medium correlation between the intelligibility of flapped tokens and their word frequency, Spearman's rho = 0.38, $p = 0.04$. Intelligibility of flapped /t/ was significantly and moderately correlated with word frequency, Spearman's rho = 0.39, $p = 0.03$. Intelligibility of flapped /d/ was also significantly and moderately correlated with word frequency, Spearman's rho = 0.41, $p = 0.02$. Self-reported experience with GA was

Table 2
Descriptive statistics for Experiment 2.

	flapped /t/	nonflapped /t/	flapped /d/	nonflapped /d/
n	30	30	30	30
Median	77.77	100	83.25	88.88
IQR	63.89	11.12	55.56	22.23
Minimum	0	44.44	11.11	22.22
Maximum	100	100	100	100

Note: IQR = Interquantile Range.

significantly and moderately related to flap intelligibility, Spearman's rho = 0.43, $p = 0.002$.

4. Discussion

Research question 1 asked to what extent does intelligibility vary according to the four types of flap. The results of Experiment 1, taken together, strongly suggest that flapped variants had a practically significant impact on intelligibility as compared to their nonflapped equivalents. First, the nasal flap was the most challenging variant for the participants, since the effect size associated with the comparison between the nasal flap and nonflapped /nt/ was very large; in addition, the CI for the effect size showed that, with 95% confidence, the true effect in the population could be a very large one. Thus we can be reasonably certain that for learners at an intermediate level and with the European language backgrounds featuring in the experiment (our population of interest), this type of flap is detrimental for intelligibility. Second, flapped /ɹt/ was likewise much less intelligible than its nonflapped counterpart. The effect size for this comparison was very large and statistically precise, as denoted by the narrow CI. For the population to which it is hoped to generalize, then, this type of flap would create critical intelligibility-related obstacles. Third, flapped /t/, the kind of flap which is the focus of the LFC, also had a large effect size, but the CI was not statistically significant since it included zero. Thus, we cannot be confident as to what the real effect could be in the population of relevance. The difference in intelligibility among the three types of flap was also investigated in Experiment 1. The post-hoc analyses indicated that the nasal flap was not different in intelligibility from flapped /ɹt/, given the trivial effect size found; the 95% CI for this effect was very wide and therefore imprecise, being statistically nonsignificant as well (it included zero). The nasal flap was considerably less intelligible than flapped /t/, with a large effect size that in the population, with 95% confidence, could be a small effect or a very large effect. This means that we can place only limited assurance as to what the effect could be in the population of interest. Finally, flapped /ɹt/ was less intelligible than flapped /t/, and this to a large extent (the effect size observed is large). We can place more confidence in this result, as the 95% CI showed that it could be a medium effect size (0.40) or a very large effect size (0.87) in the population.

The results of Experiment 2 indicated that, first, and as predicted by the LFC, nonflapped /t/ was more intelligible than flapped /t/. The effect size was large and its 95% CI indicates that it could be medium to large in the population. This means that we can have, to a certain degree, assurance that flapped /t/ interferes with intelligibility for ELF listeners with the profile described above. Second, nonflapped /d/ was in like manner more intelligible than flapped /d/, with a large effect size associated with this comparison. However, the 95% CI was very wide and therefore imprecise; the real effect in the population (with 95% confidence) could be as low as a trivial effect size 0.16 or as high as a large effect size 0.84. It could be worth conducting further research to narrow the CI (cf. Larson-Hall & Plonsky, 2015). There was only a very negligible difference ($r_b = 0.04$) in intelligibility between flapped /t/ and flapped /d/, with a 95% CI that was very wide and therefore unreliable; it is not possible to infer statistically to which degree is either flap more intelligible in the population.

Kang, Thomson, and Moran (2020) explored what specific features of accented speech make it difficult for international listeners to process speech, and found that in listening tasks the absence of flaps was rated as highly intelligible. This concurs with the overall results reported here.

Turning to the second research question, which asked to what extent is word frequency related to the intelligibility of the four types of flap, the findings are as follows.

Following Lindstromberg (2016), confidence intervals for the Spearman's correlations were not calculated, and for this reason the statistical significance (i.e., the *p*-value) of the correlations is considered in the ensuing discussion. In Experiment 1 it was found that for the three types of flap as a whole, there was a statistically significant and medium correlation between word frequency and flap intelligibility, as could be expected (more frequent words possibly being in the participants' lexical repertoire or more readily retrievable). Breaking down the results for each type of flap, there was a large and statistically significant correlation between word frequency and the intelligibility of flapped /ɹt/, a small and not statistically significant correlation between word frequency and the nasal flap, and finally a large and statistically significant correlation between word frequency and flapped /t/. As for Experiment 2, the Spearman's rho correlations between lexical frequency and intelligibility were found to be medium and statistically significant for both flapped /t/ and flapped /d/. The results of both experiments, then, suggest that the intelligibility of flapped words was, by and large, associated with their token frequency, with more frequent flapped words being easier to understand. Word frequency results are consonant with Connine and Pinnow (2006) and Tucker (2011), who found that both flapped /t/ and flapped /d/ recognition increased with higher lexical frequency.

As for the third research question, it should be noted that self-reported experience with GA did not reach statistical significance in Experiment 1, but it did in Experiment 2. Such discrepancy may be due to the small sample size in the first experiment. Thus, and expectedly, learners who are more familiar with GA found flaps less disadvantageous in intelligibility. This last result should however be taken with caution, because it was obtained through self-reporting, and self-reports may not realistically represent real language exposure and use (e.g., Colantoni, Steele, & Escudero, 2015) and this methodology can be subject to error (Flege, 2009).

4.1. Pedagogical implications (practical significance)

Jenkins (2000, 2002) included in the LFC the teaching of British-like nonflapped /t/ versus the American-based flapped /t/ because the latter can be a source of no unintelligibility among NNSs. The results of the two lab-based experiments reported here suggest that this pointer of the LFC is methodologically sound, at least in settings for learners at an intermediate level who have a European language background. While in Experiment 1 the effect size found in intelligibility between nonflapped and flapped /t/ was not statistically significant, additional data from Experiment 2 seem to fully support Jenkins' claim. The difference in intelligibility between nonflapped versus flapped /t/ in Experiment 2 was large, as per the effect size obtained, whose 95% CI ranges from a medium to a very large effect size. This indicates that when teaching pronunciation in ELF European-based contexts, like the one reproduced in this study, nonflapped /t/ should probably be emphasized, and flapped /t/ deemphasized. The difference in intelligibility between both variants is simply too vast to be ignored. These observations can confidently be extended to other variants where a flap is involved, because our results strongly suggest that the nasal flap in an ELF European language context is heavily detrimental for intelligibility, given the substantial effect size found. A very similar conclusion can be drawn for flapped /ɹt/. Finally, the intelligibility of flapped /d/ showed less conclusive evidence, with a large effect size but a CI that is too wide, which clearly calls for urgent further investigation. Another aspect to take into account is that flapped tokens which are more frequent are associated, in general and expectedly, with considerably higher intelligibility. Pronunciation instructors who need to incorporate examples of the flapped variants explored in this study could probably benefit from using more frequent words; for instance, a token like *party* may be to a large extent more intelligible than the much less frequent word *mortal*, both of which normally display the flapped /ɹt/ in accents like GA and Australian English. This frequency relationship holds relatively strongly (the correlations found are all medium to large) for all flapped variants except for the nasal flap. For instructors working in European contexts who adopt a LFC pronunciation framework, incorporating the nonflapped variants discussed in this study is probably not an arduous task,

and the potential increase in intelligibility seems to be quite large, as the effect sizes detected in both experiments indicate.

For teachers in an ESL context, if flaps need to be the focus of instruction, pronunciation methodologists such as Celce-Murcia et al. (2010) and Avery and Ehrlich (1992) remark that, at least perceptually, the GA-based flap is an important pointer in the ESL pronunciation syllabus, given its widespread occurrence. A related issue is whether this pronunciation feature is teachable (see Levis, 2018). Dalton and Seidlhofer (1994) proposed a teachability-learnability scale, whereby segments are easy to "isolate out for direct teaching" (p. 73), whereas other pronunciation components, such as intonation, are not. Matsuzawa (2006) offered the only evidence of the teachability of flapped /t/ and the nasal flap, in the context of Japanese speakers, and his results indicated that flapped /t/ can be effectively taught (the results for the nasal flap were inconclusive). In sum, pronunciation instructors based in Europe who may adopt in their teaching the LFC syllabus could probably take heed of Jenkins' proposal concerning flapped /t/ and extended it to the other flapped variants (and, at least provisionally, to flapped /d/ as well) explored in this study. Finally, learners who self-reported more familiarity with GA seemed to find the flapped variants comparably more intelligible. Thus, the quality and quantity of input seems to play a role in flap intelligibility (cf. Flege, 2009).

4.2. Limitations and further research

While this study suggested convincingly that flaps are disadvantageous for intelligibility in a lingua franca context, this research should be expanded to other situations with learners of non-European language backgrounds, ideally through replication studies. The flap is a comparatively common sound in the languages of the world (e.g., 26% frequency in Moran & McCloy, 2019; 20% in the UCLA Phonological Segment Inventory Database (UPSID)), and it occurs in major languages like Bengali, Greek, or Spanish (Ladefoged & Maddieson, 1996; Maddieson, 1984). However, it should be noted that the two experiments reported here are lab-based studies, which are per se very artificial (Hudson & Llosa, 2015), and this in turn may lessen the ecological validity of their findings and relevancy. The same concern was expressed by Jenkins (2008), requesting LFC replication studies "not in experimental conditions such as identifying items in word lists or filling gaps in passages read aloud by disembodied voices on recordings" (p. 201). In speech research, however, Xu (2010) made a convincing case for the use of laboratory studies, and in the specific context of the present study, it is not easy to imagine how to test the intelligibility of flaps without resorting to some form of experimental manipulation. Perhaps one viable methodological alternative could be the Conversation Analysis (CA) framework (e.g., O'Neal, 2021). It also needs to be acknowledged that most L2 speech research emanates from laboratory studies (for summaries see Bohn & Murray, 2007; Colantoni et al., 2015; Hansen Edwards & Zampini, 2008; Wayland, 2021). Research guidelines in this domain (e.g., Flege, 2021; Munro & Derwing, 2020) are as well framed in terms of experimental laboratory designs; therefore, it is difficult to downplay the role of variable-manipulation experiments in pronunciation research, although it would probably be wise to bear in mind its limitations, as commented above. Finally, the effect sizes reported in this study are fairly large, and as Plonsky and Oswald (2014) commented, "in the lab, researchers can often exert greater rigor and experimental control over environmental and other potentially contaminating variables, enabling greater isolation of—and thus larger—intended effects" (p. 897). More ecologically valid classroom studies where flap intelligibility can be examined are therefore clearly called for, although as noted by Sewell (2017), both natural speech data and those emanating from more controlled conditions have a place in ELF pronunciation research.

5. Conclusion

The LFC is a powerful and attractive syllabus (Levis, 2018) because it specifies certain pronunciation features as essential for ELF intelligibility. Pronunciation teachers especially have welcomed the LFC proposals enthusiastically, probably due to the widespread sentiment that English as a global language necessitates new norms for international intelligibility. However, these proposals are in need of empirical evidence. Nonflapped /t/ is a core feature of the LFC scheme, and this study sought to verify whether flapped /t/ causes unintelligibility in ELF, and extended its analysis to three other types of flap. The results suggested on the whole that flapped variants were to a large extent disadvantageous in the ELF context of European-based L1s, a relationship that is to some extent mediated first, by word frequency, with flapped words of higher lexical frequency resulting in higher intelligibility, and second, by experience with GA, whereby more global familiarity with this accent is related to higher flap intelligibility. Flaps are extensively common in accents such as GA and Australian English, and at least in perception, these sounds are given prominence in teaching schemes modeled on native performance. In lingua franca contexts of the type surveyed in the current study, nonetheless, it is recommended to resort to the non-flapped variants, given the results obtained for intelligibility. Specifically, it is suggested to practitioners that flapped /ɹt/, as in 'turtle', the nasal flap, as in 'rental', and to a lesser extent flapped /t/ as in 'fatal', are avoided in ELF contexts such as the one described in the current study. The results for flapped /d/ as in 'shady' are far less conclusive at the present time. Further, because flap intelligibility was to a moderate extent associated with word frequency and self-reported experience with GA, in instructional settings where flaps may be of relevance, likely because of their salience in many native varieties, it is recommended that higher-frequency words are considered as potential teaching targets and that the degree of exposure of leaners to flapping accents (such as GA) is taken into account.

CRediT authorship contribution statement

Darío Barrera-Pardo: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The author is appreciative of very helpful comments from two anonymous reviewers.

Appendix A

Target words of Experiment 1

1. Nasal flaps

1. interview
2. enter
3. dentist
4. plenty
5. rental
6. winter
7. wanted
8. county

9. center
10. granted
11. painted
12. twenty

2. flapped /ɹt/

1. party
2. mortal
3. turtle
4. started
5. thirty
6. charter
7. forty
8. sorting
9. dirty
10. shorty
11. myrtle
12. fertile

3. flapped /t/

1. united
2. beauty
3. matter
4. writer
5. native
6. notice
7. mighty
8. pity
9. hotter
10. rattle
11. beetle
12. kettle

Appendix B

Target words of Experiment 2

1. Flapped /t/

1. battle
2. fatter
3. total
4. daughter
5. hater
6. city
7. rattle
8. boating
9. petting
10. meeting
11. better
12. fighter
13. little
14. hotter
15. beetle
16. bitter
17. butter
18. voter
19. shooter
20. bottle
21. fatal
22. settle
23. cattle
24. title
25. pity
26. chatting
27. betting
28. biting
29. dating
30. liter

2. Flapped /d/

1. dreading
2. shading
3. goodies
4. gliding
5. speedy
6. shady
7. tidy
8. breeding
9. studying
10. lady
11. leading
12. ready
13. reading
14. body
15. study
16. woody
17. hiding
18. loading
19. feeding
20. faded
21. traded
22. bleeding
23. pedal
24. flooded
25. moody
26. needed
27. heading
28. buddy
29. teddy
30. daddy

References

Avery, P., & Ehrlich, S. (1992). *Teaching American English pronunciation*. Oxford: Oxford University Press.

Barrass, J., Baffoe-Djan, J. B., Rose, H., & Boggs, J. (2020). Intelligibility and comprehensibility of Korean English speakers' phonological features in lingua franca listening contexts. *Journal of Asia TEFL, 17*(1), 1–19. https://doi.org/10.18823/asiatefl.2020.17.1.8.124

Bohn, O.-S., & Munro, M. J. (Eds.). (2007). *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/lllt.17.

Brown, J. D. (1988). *Understanding research in second language learning: A teacher's guide to statistics and research design*. New York: Cambridge University Press.

Byrd, D. (1993). 54,000 American stops. *UCLA Working Papers in Phonetics, 83*, 97–116.

Catford, J. C. (1987). Phonetics and the teaching of pronunciation. In J. Morley (Ed.), *Current perspectives on pronunciation* (pp. 87–100). Washington, D.C.: TESOL.

Celce-Murcia, M., Brinton, D., Goodwin, J., & Giner, B. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). New York: Cambridge University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum.

Colantoni, L., Steele, J., & Escudero, P. (2015). *Second language speech: Theory and practice*. Cambridge: Cambridge University Press.

Connine, C. M., & Pinnow, E. (2006). Phonological variation in spoken word recognition: Episodes and abstractions. *The Linguistic Review, 23*, 235–245. https://doi.org/10.1515/TLR.2006.009

Cox, F., & Fletcher, J. (2017). *Australian English: Pronunciation and transcription* (2nd ed.). Melbourne: Cambridge University Press.

Cummings, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.

Dalton, C., & Seidlhofer, B. (1994). *Pronunciation*. Oxford: Oxford University Press.

Dauer, R. M. (2005). The Lingua Franca Core: A new model for pronunciation instruction? *Tesol Quarterly, 39*(3), 543–550. https://doi.org/10.2307/3588494

Davies, M. (2019). Corpus of contemporary American English. Available online at https://www.english-corpora.org/coca/.
Demir, Y., & Kartal, G. (2022). Mapping research on L2 pronunciation: A bibliometric analysis. *Studies in Second Language Acquisition*. https://doi.org/10.1017/S0272263121000966
Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *Tesol Quarterly, 39*(3), 379–397. https://doi.org/10.2307/3588486
Deterding, D. (2013). *Misunderstandings in English as a lingua franca: An analysis of ELF interactions in South-East Asia*. Berlin: Mouton de Gruyter.
Eddington, D., & Elzinga, D. (2008). The phonetic context of American English flapping: Quantitative evidence. *Language And Speech, 51*(3), 245–266.
Egbert, J., & LaFlair, G. T. (2018). Statistics for categorical, nonparametric, and distribution-free data. In A. Phakiti, P. I. De Costa, L. Plonsky, & S. Starfield (Eds.), *The Palgrave handbook of applied linguistics research methodology* (pp. 523–539). New York: Palgrave.
Flege, J. E. (2009). Give input a chance. In T. Piske, & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 175–190). Bristol: Multilingual Matters.
Flege, J. E. (2021). New methods for second language (L2) speech research. In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 119–156). Cambridge: Cambridge University Press.
Gadiner, I. A. (2019). Intelligibility of an L2 variety in ELF interactions: Brunei English and vowel pronunciation. *Journal of Second Language Pronunciation, 5*(2), 247–268. https://doi.org/10.1075/jslp.17049.gar
Graddol, D. (2006). *English next: Why global English may mean the end of 'English as a foreign language'*. London: British Council. Available at https://www.teachingenglish.org.uk/article/english-next.
Hansen Edwards, J. G., & Zampini, M. L. (Eds.). (2008). *Phonology and second language acquisition*. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/sibil.36.
Hudson, T., & Llosa, L. (2015). Design issues and inference in experimental L2 research. *Language Learning, 65*(Suppl. 1), 76–96. https://doi.org/10.1111/lang.12113
Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.
Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics, 23*, 83–103. https://doi.org/10.1093/applin/23.1.83
Jenkins, J. (2007). *English as a lingua franca: Attitude and identity*. Oxford: Oxford University Press.
Jenkins, J. (2008). Misinterpretation, bias, and resistance to change: The case of the Lingua Franca Core. In K. Dziubalska-Kolacyk, & J. Przedlacka (Eds.), *English pronunciation models: A changing scene* (2nd ed., pp. 77–98). Frankfurt am Main: Peter Lang.
Jenkins, J. (2009). *World Englishes: A resource book for students* (2nd. ed.). New York: Routledge.
Jeong, H., & Thorén, B. (2018). Intelligibility of the alveolar [s] replacing the interdental [θ] in English words. In Å. Abelin, & Y. Nagano-Madsen (Eds.), *Proceedings FONETIK 2018* (pp. 39–41). University of Gothenburg. The XXXth Swedish Phonetics Conference Gothenburg, June 7–8 https://www.bossethoren.se/bosse_Jeong_Thoren_fon2018.pdf.
Jeong, H., Thorén, B., & Othman, J. (2020). Effect of altering three phonetic features on intelligibility of English as a lingua franca: A Malaysian speaker and Swedish listeners. *Asian Englishes, 22*(1), 2–19. https://doi.org/10.1080/13488678.2018.1536817
Jurado-Bravo, M. A. (2018). Vowel quality and vowel length in English as a lingua franca in Spain. *Miscelanea: A Journal of English and American Studies, 57*, 13–34. https://doi.org/10.26754/ojs_misc/mj.20186310
Kang, O., Thomson, R. I., & Moran, M. (2020). Which features of accent affect understanding? Exploring the intelligibility threshold of diverse accent varieties. *Applied Linguistics, 41*(4), 453–480. https://doi.org/10.1093/applin/amy053
Kennedy, S. (2012). When non-native speakers misunderstand each other: Identifying important aspects of pronunciation. *Contact, 38*(2), 49–62.
Kiczkowiak, M. (2020). Seven principles for writing materials for English as a lingua franca. *ELT Journal, 74*(1), 1–9. https://doi.org/10.1093/elt/ccz042
Ladefoged, P., & Maddieson, I. (1996). *Sounds of the world's languages*. Oxford: Blackwell.
Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluative reactions to spoken languages. *Journal of Abnormal and Social Psychology, 60*, 44–51.
Lane, L. (2010). *Tips for teaching pronunciation: A practical approach*. White Plains, NY: Pearson Education.
Larson-Hall, J. (2015). *A guide to doing statistics in second language research using SPSS and R* (2nd ed.). London: Routledge.
Larson-Hall, J., & Plonsky, L. (2015). Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning, 65*, 127–159. https://doi.org/10.1111/lang.12115. Supp. 1.
Levis, J. M. (2016). The interaction of research and pedagogy. *Journal of Second Language Pronunciation, 2*(1), 1–7. https://doi.org/10.1075/jslp.2.1.001lev
Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge: Cambridge University Press.
Lewis, C., & Deterding, D. (2018). Word stress and pronunciation teaching in English as a lingua franca contexts. *The CATESOL Journal, 30*(1), 161–176.
Lindstromberg, S. (2016). Guidelines, recommendations, and supplementary discussion. *Language Teaching Research*. https://doi.org/10.1177/1362168816651895
Loerts, H., Lowie, W., & Seton, B. (2020). *Essential statistics for applied linguistics (using R or JASP)* (2nd ed.). London: Red Globe Press.
Luchini, P. L., & Kennedy, S. (2013). Exploring sources of phonological unintelligibility in spontaneous speech. *International Journal of English Language and Literature, 4*(3), 79–88. https://doi.org/10.5897/IJEL12.049
Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.
Major, R. C. (2008). Transfer in second language phonology: A review. In J. G. Hansen Edwards, & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 63–94). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/sibil.36.05maj.
Matsuzawa, T. (2006). Comprehension of English reduced forms by Japanese business people and the effectiveness of instruction. In J. D. Brown, & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers* (pp. 59–66). Honolulu, HI: University of Hawai'i, National Foreign Language Resource Center.
McCrocklin, S. (2012). The role of word stress in English as a lingua franca. In J. Levis, & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in second language Learning and teaching conference, Sept. 2011* (pp. 249–256). Ames, IA: Iowa State University.
Moran, S., & McCloy, D. (2019). *PHOIBLE 2.0*. Jena: Max Planck Institute for the Science of Human History. Available online at http://phoible.org. (Accessed 7 February 2022).
Moyer, A. (2009). Input as a critical mean to an end: Quantity and quality of experience in L2 phonological attainment. In T. Piske, & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 159–174). Bristol: Multilingual Matters.
Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, an intelligibility in the speech of second language learners. *Language Learning, 45*(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x
Munro, M. J., & Derwing, T. M. (2015). A prospectus for pronunciation research in the 21st century: A point of view. *Journal of Second Language Pronunciation, 1*(1), 11–42. https://doi.org/10.1075/jslp.1.1.01mun
Munro, M. J., & Derwing, T. M. (2020). Collecting data in L2 pronunciation research. September 2019. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (Eds.), *Proceedings of the 11th pronunciation in second language learning and teaching conference* (pp. 8–18). Ames, IA: Northern Arizona University. Iowa State University.
Munro, M. J., Derwing, T. M., & Holtby, A. K. (2012). Evaluating individual variability in foreign accent comprehension. In J. Levis, & K. LeVelle (Eds.), *Proceedings of the 3rd pronunciation in second language learning and teaching conference, Sept. 2011* (pp. 233–239). Ames, IA: Iowa State University.
Norris, J. M., Plonsky, L., Ross, S. J., & Schoonen, R. (2015). Guidelines for reporting quantitative methods and results in primary research. *Language Learning, 65*(2), 470–476. https://doi.org/10.1111/lang.12104
O'Neal, G. (2015). ELF intelligibility: The vowel quality factor. *Journal of English As a Lingua Franca, 4*(2), 347–358. https://doi.org/10.1515/jelf-2015-0026
O'Neal, G. (2019). The accommodation of intelligible segmental pronunciation: Segmental repairs and adjustments in English as a Lingua Franca interactions. *Journal of Second Language Pronunciation, 5*(1), 119–138. https://doi.org/10.1075/jslp.17002.one
O'Neal, G. (2020). Does an ELF phonology exist? *Asian Englishes, 22*(3), 282–296. https://doi.org/10.1080/13488678.2019.1706135
O'Neal, G. (2021). What is the effect of successive segmental repair on the mutual intelligibility of EFL users? *System, 103*, Article 102683. https://doi.org/10.1016/j.system.2021.102683
Osimk, R. (2009). Decoding sounds: An experimental approach to intelligibility in ELF. *Vienna English Working Papers, 18*, 64–89.
Patterson, D., & Connine, C. M. (2001). Variant frequency in flap production: A corpus analysis of variant frequency in American English flap production. *Phonetica, 58*(4), 254–275. https://doi.org/10.1159/000046178
Picard, M. (2007). On teaching the pronunciation of subphonemic segments in English. *SKY Journal of Linguistics, 20*, 331–344.
Picard, M. (2010). *Flapping in North American English: A complete guide to ESL teachers*. Saarbrücken: VDM Verlag Dr. Müller.
Pickering, L., & Huang, M. (2022). Teaching pronunciation in the context of multiple varieties of English. In J. M. Levis, T. M. Derwing, & S. Sonsaat-Hegelheimer (Eds.), *Second language pronunciation: Bridging the gap betweeen research and teaching* (pp. 273–292). Oxford: Wiley Blackwell.
Pitt, M. A., Dilley, L., & Tat, M. (2011). Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics, 39*(3), 304–311. https://doi.org/10.1016/j.wocn.2010.07.004
Pitzl, M. L. (2018). Transient international groups (TIGs): Exploring the group and development dimension of ELF. *Journal of English As a Lingua Franca, 7*(1), 25–58. https://doi.org/10.1515/jelf-2018-0002
Plonsky, L. (2015). Statistical power, *p* values, descriptive statistics, and effect sizes: A "back-to-basics" approach to advancing quantitative methods in L2 research. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 23–45). New York: Routledge.
Plonsky, L. (2021). Quantitative research methods and the reform movement in applied linguistics. In H. Mohebbi, & C. Coombe (Eds.), *Research questions in language education and applied linguistics* (pp. 749–753). New York: Springer. https://doi.org/10.1007/978-3-030-79143-8_130.
Plonsky, L., & Oswald, F. L. (2014). How big is 'big'? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912. https://doi.org/10.1111/lang.12079
Rahimi, M., & Ruzrokh, S. (2016). The impact of teaching Lingua Franca Core on English as a foreign language learners' intelligibility and attitudes towards English pronunciation. *Asian Englishes, 18*(2), 141–156. https://doi.org/10.1080/13488678.2016.1173466
Rau, D., Chang, H.-H. A., & Tarone, E. E. (2009). Think or sink: Chinese learners' acquisition of the English voiceless interdental fricative. *Language Learning, 59*(3), 581–621. https://doi.org/10.1111/j.1467-9922.2009.00518.x

Schleef, E. (2013). Glottal replacement of /t/ in two British capitals: Effects of word frequency and morphological compositionality. *Language Variation and Change, 25* (2), 201–223. https://doi.org/10.1017/S0954394513000094

Sewell, A. (2017). Functional load revisited: Reinterpreting the findings of 'lingua franca' intelligibility studies. *Journal of Second Language Pronunciation, 3*(1), 57–79. https://doi.org/10.1075/jslp.3.1.03sew

Shockey, L. (2003). *Sound patterns of spoken English*. Oxford: Blackwell.

Suntornsawet, J. (2019). Problematic phonological features of foreign accented English pronunciation as threats to international intelligibility: Thai EIL pronunciation core. *Journal of English as an International Language, 14*(2), 72–93. https://files.eric.ed.gov/fulltext/EJ1244216.pdf.

Szpyra-Kozłowska, J. (2008). Lingua Franca Core, phonetic universals and the Polish context. In K. Dziubalska-Kołaczyk, & J. Przedlacka (Eds.), *English pronunciation models: A changing scene* (2nd ed., pp. 151–176). Peter Lang: Bern.

Szpyra-Kozłowska, J. (2015). *Pronunciation in EFL instruction: A research-based approach*. Bristol: Multilingual Matters.

Tanner, M., & Henrichsen, L. (2022). Pronunciation in varied teaching and learning contexts. In J. M. Levis, T. M. Derwing, & S. Sonsaat-Hegelheimer (Eds.), *Second language pronunciation: Bridging the gap betweeen research and teaching* (pp. 215–234). Oxford: Wiley Blackwell.

Thir, V. (2020). 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

Trofimovich, P., Collins, L., Cardoso, W., White, J., & Horst, M. (2012). A frequency-based approach to L2 phonological learning: Teacher input and student output in an intensive ESL context. *Tesol Quarterly, 46*(1), 176–186. https://doi.org/10.1002/tesq.9

Trudgill, P. (2008). Native speaker segmental phonological models and the Lingua Franca Core. In K. Dziubalska-Kołacyk, & J. Przedlacka J (Eds.), *English pronunciation models: A changing scene* (2nd. ed., pp. 77–98). Peter Lang: Bern.

Tucker, B. V. (2011). The effect of reduction on the processing of flaps and /g/ in isolated words. *Journal of Phonetics, 39*(3), 312–318. https://doi.org/10.1016/j.wocn.2010.12.001

UCLA. Phonological segment inventory Database (UPSID). Available from http://web.phonetik.uni-frankfurt.de/upsid_info.html.

Walker, R. (2010). *Teaching the pronunciation of English as a lingua franca*. Oxford: Oxford University Press.

Wayland, R. (Ed.). (2021). *Second language speech learning: Theoretical and empirical progress*. Cambridge: Cambridge University Press.

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

Xu, Y. (2010). In defense of lab speech. *Journal of Phonetics, 38*(3), 329–336. https://doi.org/10.1016/j.wocn.2010.04.003

Zhang, L. (2013). International intelligibility of Chinese-accented English. In Y. Bayyurt, & S. Akcan (Eds.), *ELF5, proceedings of the fifth international conference of English as a Lingua Franca* (pp. 327–334). May 24-26 2012, Istanbul.

Zoghbor, W. S. (2018). Teaching English pronunciation to multi-dialect first language learners: The revival of the Lingua Franca Core (LFC). *System, 78*, 1–14. https://doi.org/10.1016/j.system.2018.06.008