

# Acoustic (In)validity of Breton TTS: A Comparative Vowel Space Analysis with Training corpus and Native Production

Seongwoo KANG

University of Western Brittany, CRBC

*OLMi, Plouzané, Brittany*

5 May 2026

## Abstract

This paper evaluates the acoustic accuracy of a Breton Text-to-Speech (TTS) system, *Sintezenn*, by comparing its vowel space to both its TTS training data and traditional Breton speakers from the north-eastern part of Breton-speaking Brittany. The paper uses vowel space analysis with *Visible Vowels*[22] and Bayesian statistical models to determine if the Breton TTS successfully replicates the vowel space (defined by formants) of the TTS training corpus used to build it. Results indicate high engineering fidelity, meaning the technology mirrors its training recordings with great precision. However, a slight sociophonetic gap exists between the TTS output and traditional Breton speakers from the Tregor region, particularly regarding specific vowel height and backness. While the Breton TTS is deemed highly intelligible for educational and public use, our paper highlights distinct differences in how certain vowels, like /ø/ and /a/, are realised compared to traditional Breton speech.

## 1 Introduction

With the rapid advancement of artificial intelligence, minority languages are increasingly entering the generative AI era, and Breton is no exception [1, 2]. Recently, Breton has been integrated into ChatGPT and Google Translate. Furthermore, *Sintezenn*<sup>1</sup>, a Breton Text-to-Speech (TTS) service developed by IRISA and funded by the Brittany Region and the Public Office of the Bre-

ton Language (OPLB), is now publicly available. The potential applications for Breton TTS are extensive: generated audio can be employed in educational settings as mentioned explicitly by the OPLB<sup>2</sup>, public services (such as automated announcements in local trains), and linguistic research. For scholars in phonetics and psycholinguistics, TTS facilitates the generation of stimuli that were previously difficult to produce consistently.

### 1.1 Motivation

TTS can be employed in various domains of daily life, as previously noted. It serves multiple roles in our society: automated commercial announcements, applications in Large Language Models (LLMs) such as ChatGPT and Gemini for conversation, and other similar functions. However, as phonologists and phoneticians, we primarily focus on its potential in *pedagogical enterprise*. The number of traditional native speakers has been decreasing over the years. In a survey conducted in 2018[3], the estimated number of speakers was 207,000, and approximately 78 percent of speakers were older than 60 years old (individuals assumed to be traditional speakers). However, in a survey conducted in 2024[4], the estimated number of speakers dropped drastically to 107,000 and only 60 percent of speakers are older than 60 years old. In this context, where generational transmission is interrupted [5, p. 3], language courses in schools or for adults have become more common methods

<sup>1</sup><https://niverel.breizh.bzh/fr/sintezenn/>

<sup>2</sup>"Application pédagogiques ou ludiques"

of Breton acquisition than traditional transmission through family or neighbors.

Our motivation for this paper stems from this sociolinguistic situation regarding Breton transmission and learning. If Breton TTS is designed for pedagogical purposes, which variety of Breton does it use? Could it be an optimal candidate for Breton learning (more precisely, traditional Breton learning)?

## 1.2 Review Breton TTS evaluation

In this subsection, we will conduct a brief review of Breton TTS evaluation, focusing especially on Guennec et al. [6], a conference paper illustrating the training corpus, the model used to develop Breton TTS, and their *evaluations*, as well as Kang [7], a conference paper which phonologically evaluated the prosody of Breton TTS.

### 1.2.1 Guennec et al. 2022

In their paper, the authors (who are also the developers of the Breton TTS) present the training corpus used to develop their TTS model, as well as the system used to develop the TTS. They used an end-to-end system to generate a sound file from a text file. More precisely, they convert text (written in *peurunvan* orthography) to phonemes; then, *Tacotron 2* converts those phonemes to acoustic values based on its training corpus, and finally, the *ParallelWaveGan* vocoder converts these acoustic values to an actual waveform (a sound file).

For their training corpus, they cast two (one male and one female) native speakers, both from the Tregor region (the northeastern part of Breton-speaking Brittany), selected based on criteria such as speech intelligibility, dynamism, voice quality, and their flexibility to adapt their pronunciation to *peurunvan*. The male speaker was 48 years old at the time; he has a traditional background and is considered a *linguistic expert* for the project. The female speaker has a more standard background than a traditional one; she was a long-time anchorwoman on Breton television.

For the evaluation of generated speech, they contacted *experienced Breton speaker(s)* to evaluate the samples on a scale of 1 (bad) to 5 (excellent) (using the Mean Opinion Score, henceforth MOS). The male TTS voice was rated 3.93 for

grapheme-based generated speech files and 3.73 for phoneme-based generated speech files. The female TTS voice was rated 3.74 for grapheme-based and 3.84 for phoneme-based files. These scores can be interpreted, as noted by the authors [6, p. 6], as representing *good quality*.

We can identify some lack of clarity in their approach, regarding both the speaker selection criteria and the evaluation method. Firstly, they did not specify what they meant by *speech intelligibility*, *dynamism*, and *voice quality*. They provided no definition as to whether these criteria are based on psychological and perceptual aspects of speech or on the actual acoustic aspects of Breton speech, which remain understudied.

Secondly, the MOS evaluation method is highly subjective and behavioural. In addition to the fact that they did not specify the profiles or the number of their *experienced Breton speaker(s)*, the linguistic (phonetic/phonological) validity of this evaluation method is also questionable.

### 1.2.2 Kang 2025

In the paper titled *Évaluation comparative de l'intonation de la synthèse vocale du breton et de la voix dans la matière pédagogique du breton*, which shares the same motivations as the present study, we calculated the phonological distance of prosodic structures between traditional speakers (from Central Brittany) and the speech generated by Breton TTS. By prosodic structure, we refer to pitch accent distribution [8], stress in negative sentences, and sentence stress and deaccenting [9]. To compare the prosodic structures of traditional speakers and Breton TTS, we employed the metrical grid [10] and Hamming distance [11], which are familiar to linguists (phonologists and dialectologists).

Results showed an overall phonological similarity (mean value) of 92 percent between the prosodic structure of traditional speakers and the prosodic structure of Breton TTS. An 8 percent difference is not a large one in terms of phonetic distance [12, p. 79], but in terms of structure, 8 percent is not negligible, given that a structural (or phonological) difference can also be referred to as a difference in stress position, the existence of a pitch accent, etc.

Certainly, linguistic validity (representing a

distance-based similarity or passing a statistical test) does not necessarily mean that a linguistically valid model is actually natural to Breton speakers. However, in a sociophonetic study on *nativelikeness* or *fluency* judgments of L2 English speakers, Mairano et al. [13] found that acoustic/phonetic-based evaluation methods (such as *Linear Discriminant Analysis* and *Pillai Scores*) actually correspond well to native judgments. This finding suggests that acoustic/phonetic-based statistical methods could be used as a perception model to evaluate the naturalness of a given speech sample.

## 2 Research Questions

From our preliminary reflections, we have formulated two objectives for evaluating TTS-generated sentences. Firstly, from an engineering perspective, it is important to see if the TTS output corresponds well to the training corpus<sup>3</sup>. This can be correctly evaluated by comparing the TTS-generated speech with the training corpus. Secondly, from a sociophonetic perspective, it is important to determine whether the acoustic features of TTS-generated speech correspond to actual speech—more precisely, to the acoustic features of traditional speakers. Since one example of the utilisation of Breton TTS is pedagogical use, we could eventually question which variety of Breton is transferred to schools using Breton TTS: would it be closer to that of traditional speakers, and from which region? To evaluate the acoustic features of Breton TTS, we chose *vowel space* as our first investigation. Breton phonetics is understudied using contemporary phonetic and statistical methods (except for [14, 15, 16]), regarding both segmental and prosodic features. Vowel space refers to the range of possible vowel sounds, measured by formants ( $F1$  representing tongue height,  $F2$  representing tongue backness, etc.). We chose vowel space because vowels are one of the segmental features frequently used to evaluate L2 speech [13, 17].

To address our two objectives from both engineering and sociophonetic perspectives, we have formulated two research questions, namely:

- (1) To what extent does the vowel space of the Breton TTS output align with its training corpus? (engineering perspective)
- (2) To what extent is the vowel space of the Breton TTS (and its training corpus) acoustically comparable to that of traditional speakers? (sociophonetic perspective)

Taken together, these two research questions will allow us to examine the engineering fidelity of Breton TTS and its sociophonetic representativeness regarding traditional Breton speakers, which could have methodological consequences for TTS development in minority languages.

## 3 Corpora

To answer these two specific research questions, we selected three corpora: one representing traditional Breton speakers, one representing the TTS training corpus, and one representing the TTS-generated corpus. In the following subsections, we will justify our dialectal choice for the traditional Breton corpus, as well as the sentences selected for TTS generation.

### 3.1 Traditional Breton corpus

For the traditional Breton corpus, we deliberately used the *Brezhoneg Bro Vear* (Breton of Bégard county) corpus<sup>4</sup>. *Brezhoneg Bro Vear* is situated in the Tregor region (the north-eastern part of Breton-speaking Brittany); it is, indeed, a central Tregor dialect [18, p. 6]. The reasons why we deliberately chose a north-eastern variety to represent *traditional Breton* can be summarised by the following three facts. Firstly, for pragmatic reasons, we selected *Brezhoneg Bro Vear* as it is the most accessible corpus with segmented (sentence-by-sentence or word-by-word) recordings accompanied by graphemic transcriptions (and, in many cases, phonetic transcriptions). Secondly, Bégard county is one of the areas where Breton is relatively more active than in others [4, p. 18]; statistically, therefore, one would expect to find more traditional speakers in this region. Finally, the

<sup>3</sup>This idea was suggested by an anonymous reviewer, whom we sincerely thank for this insight.

<sup>4</sup><https://www.brezhonegbrovear.bzh/homepage>; we sincerely thank Tangi Yekel for having given us authorisation to use his recordings

Tregor region is where the two TTS training corpus speakers originate [6, p. 4]. If one were to conduct a linguistic comparison between given groups, it is more relevant to select a group with the same geographical background (especially if other conditions cannot be met).

For the vowel phonemes analysed, we followed the phonetic distinction of Yekel [18] and the phonemic distinction<sup>5</sup> of Le Dû [19]. The analysed vowel phonemes are as follows: /i/, /y/, /e/, /ø/, /ɛ/, /a/, /u/, /o/, /ɔ/. Due to corpus limitations (specifically regarding the TTS training corpus rather than the traditional Breton corpus), we only compare vowels in stressed positions.

We selected five male and five female speakers, all of them from Bégard county, as defined by Yekel [18, p. 5], born between the 1930s and the 1950s. All of our phonemic vowel annotations were performed in Praat[20], marking only target vowels within the *TextGrid* interface. We then used the Python library *Parselmouth*[21] to extract the formant values ( $F1$  and  $F2$ ). We initially aimed to collect ten items for each vowel phoneme, which would result in an expected 900 vowel items; however, some vowels were particularly rare (specifically /u/, /o/, and /ɔ/), so we only obtained 693 vowel items in total.

### 3.2 TTS training corpus

For the TTS training, we utilised the entire TTS training corpus available in the OPLB’s audio corpus<sup>6</sup>, comprising two speakers. As mentioned in Guennec et al. [6, p. 4], both the male and female speakers come from the Tregor region. They exhibit a largely Tregor/standard pronunciation, but one could expect that they deliberately adapted to the so-called *standard pronunciation* given that they were reading sentences written in *peurunvan* orthography, which is widely used in education for both schools and adult language learning centres.

We adopted the same phonemic vowel inventory as the traditional Breton corpus to select sentences

<sup>5</sup>We firstly followed Yekel’s transcription closely and conducted a preliminary acoustic analysis of vowel formants; we found that /ø/ and /œ/ have converged and cannot be distinguished. Therefore, for these vowel phonemes, we treat them all phonologically as /ø/, which was already noted by Jean Le Dû in his study of Plougrescant Breton.

<sup>6</sup><https://github.com/Ofis-publik-ar-brezhoneg/audio-breton-corpus>

carrying the target vowels. To avoid boundary tone effects, we generally ignored words found at the end of sentences. In sum, we selected 46 sentences from the OPLB’s audio corpus. Ideally, we would have gathered 180 vowel items (10 items for each vowel phoneme), but due to constraints regarding position and coarticulation, only 163 vowel items were collected (with only 5 items gathered for /ø/). A list of the selected sentences is available in the annex of this paper.

### 3.3 TTS corpus

We employed the same sentences as those in the TTS training corpus to generate and analyse the target vowels. The same conditions were applied as in the TTS training corpus. We selected two voices (*Paol* and *Aziliz*), which are directly based on the two speakers from our TTS training corpus. We collected 169 vowel items.

## 4 Method

To address our two research questions, we will first plot the vowel space of each source using *Visible Vowels* [22]. Once vowel formant values are normalised using Lobanov’s normalisation [23], we will qualitatively examine the differences between the corpora (including interpersonal variation within each corpus), as well as differences regarding specific vowels and the direction of these shifts (vowel height and backness).

Statistical tools, as demonstrated by Mairano et al. [13], provide a robust method for objectively verifying qualitatively observed differences. We will train two Bayesian models [24] to investigate acoustic differences between our corpora. A Bayesian mixed-effects model, implemented using the Python libraries *PyMC* [25], *Bambi* [26] and *ArviZ* [27], will be employed to analyse acoustic differences between the traditional Breton corpus and the TTS output. In this analysis, the acoustic cues (the dependent variables) will reflect the interaction between register (traditional vs. TTS training vs. TTS) and vowel type, while accounting for interpersonal variation. This relationship can be represented by the following formula:

$$\text{Formants} \sim \text{Register} \times \text{Vowel\_types} + (1 | \text{Speaker})$$

To provide a tangible comparison between the three registers, we will compare the posterior distributions (mean values and the *Highest Density Interval (HDI)*) of each vowel produced by speakers from each register. This will provide a statistically faithful representation of the *differences* between the registers.

## 5 results

In this section, we will visualise the vowel formant values extracted from each corpus. This will give a preliminary impression of whether there are differences between the corpora. Then, a statistical analysis will provide insight into whether these differences are real or merely a visual illusion. Furthermore, a Bayesian statistical analysis will permit us to see which vowels tend to be more stable and how they diverge from one corpus to another.

### 5.1 Visible Vowels

*Visible Vowels*<sup>8</sup> offers a wide range of functions. Our analyses of the vowel space will begin with an overall visualisation of each corpus. Then, a precise visualisation of the vowel range will permit us to compare the realisation of each vowel, which will allow us to determine whether the realisation of a given vowel is consistent.

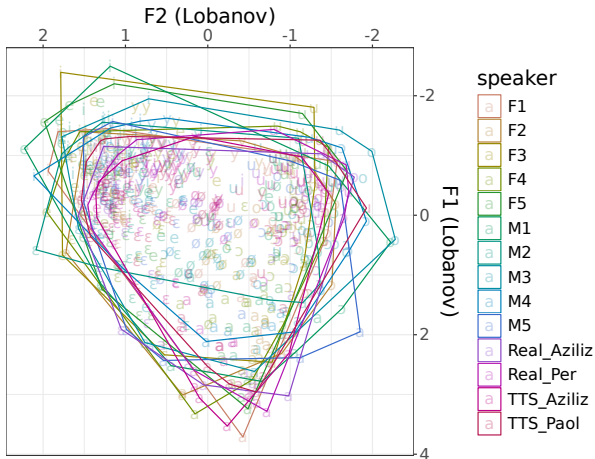


Figure 1: Vowel space of Traditional Breton, TTS training and TTS corpus

We can see that the normalised vowel space of the traditional Breton corpus is not uniform, and

<sup>7</sup>Register means audio source.

<sup>8</sup><https://visiblevowels.org/>

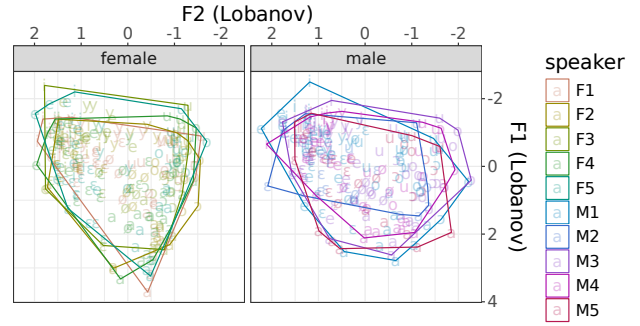


Figure 2: Vowel space of Traditional Breton

varies substantially in terms of sex. Figure 2 shows that female speakers produce a more lowered /a/ (*i.e.*, higher  $F1$  values) in the CV: condition. Furthermore, other mid and high back vowels (/u/, /o/ and /ɔ/) are pronounced with a slightly more fronted tongue position than those of male speakers (*i.e.*, higher  $F2$  values).

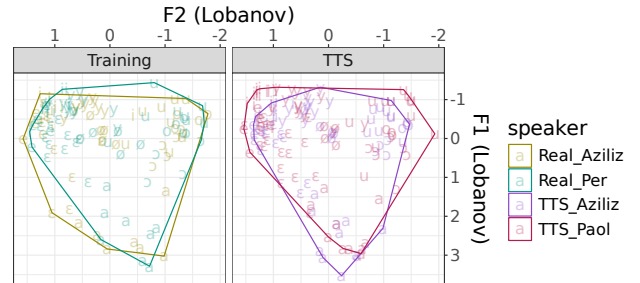


Figure 3: Vowel space of TTS training and TTS corpus

Comparing the vowel space of the TTS training corpus and the TTS corpus, we observe few differences between these two sources. The one notable difference is the realisation of /a/ in the CVC condition, where this vowel is pronounced more centrally (*i.e.*, slightly lower  $F1$  and  $F2$  values) in the TTS corpus than in the TTS training corpus. Notably, in both corpora, we can hardly perceive differences between these two speakers. This suggests that the normalisation method [23] successfully re-

moved biological differences between the two sexes, and also that there are no other remarkable differences (dialectal or sociolectal) between them. The non-normalised vowel space shows that, while female speakers have a much wider vowel space (since females have shorter vocal tracts, higher vocal fold vibration frequency is expected), the shape of the hulls is quite similar (Figure 4).

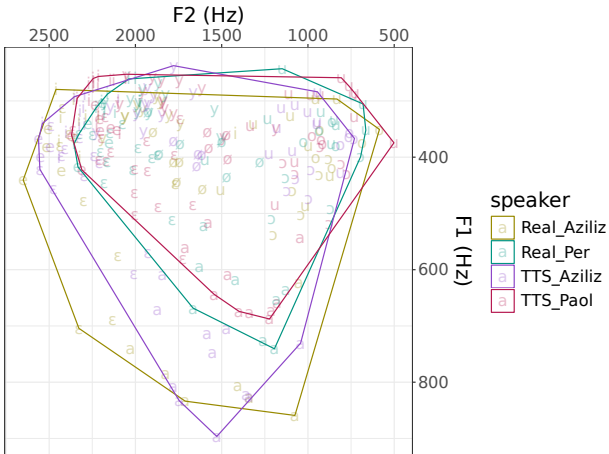


Figure 4: Vowel space of TTS training and TTS corpus (non-normalised)

This also holds true when the vowel spaces of all speakers are overlapped according to sex. We can see that traditional speakers show a slightly different shape (with male speakers having a more rectangular form while female speakers have a more triangular form). However, the vowel spaces of male speakers in the TTS training and TTS corpora seem to be more similar to the female ones (regardless of register).

Moving to vowel range and distribution, we can observe notable differences in certain vowels between traditional Breton and other registers. In the following paragraphs, we will only analyse differences across registers, regardless of sex and interpersonal differences, as register differences are the main scope of our paper.

We can observe that there are few differences in terms of vowel range. Only high back vowels (/u/ and /o/) seem to differ. However, as seen in Figure 7, the vowel range of /u/ and /o/ in both TTS training and TTS corpora is almost identical, with /u/ and /o/ being almost merged.

While there are few differences between the TTS training and TTS corpora, we can observe some

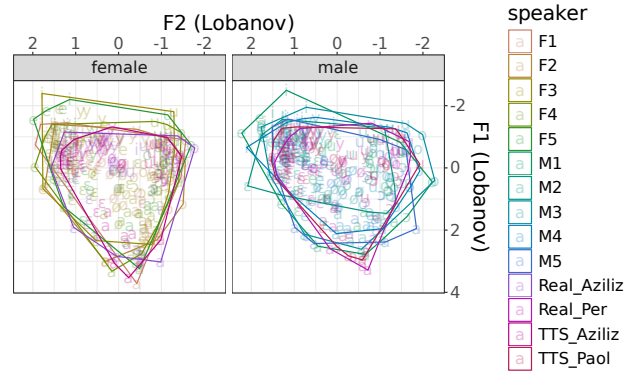


Figure 5: Vowel space overlapped according to speakers' sex)

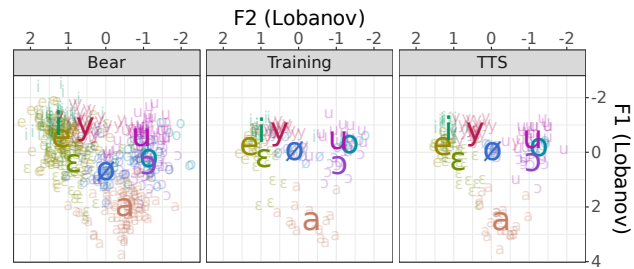


Figure 6: Centroid vowels of each register

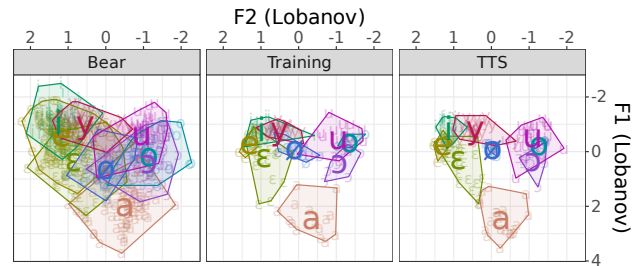


Figure 7: Vowel range of each register

notable differences between the traditional Breton corpus and the others. In addition to the fact that traditional Breton speakers have a wider vowel range, firstly, we can note that /e/ is pronounced with a slightly more fronted tongue position (higher  $F2$  values) in the TTS training and TTS corpora; their /e/ is more fronted than /i/. Secondly, we can observe that /u/ in the TTS training and TTS corpora is pronounced with the tongue positioned lower (higher  $F1$  values) than for traditional Breton speakers, which explains why /u/ and /o/ are clearly merging. Thirdly, /a/ in the TTS training and TTS corpora is pronounced with a more fronted tongue position. Furthermore, it appears that the difference between /a/ found in CV: and /a/ in CVC is well represented in all three corpora; /a/ in the CVC condition is usually pronounced with a more fronted tongue position and /a/ in CV: is pronounced with a more backed tongue position, although traditional Breton speakers seem to pronounce /a/ in CV: with an even more backed tongue position. Finally, the most notable difference between traditional Breton speakers and the other registers is /ø/. While /ø/ in the TTS training and TTS corpora is more of a high front vowel closer to /e/, traditional Breton speakers' /ø/ seems to be very central, much closer to /ɛ/.

We impressionistically observed some convergences and divergences between our three corpora in terms of vowel space. Clearly, the TTS training and TTS corpora seem to have the same realisation of vowels. Also, comparing the traditional Breton corpus and the other registers, we found that globally the shapes of the vowel space could be regarded as similar, but the distribution of some vowels may differ.

## 5.2 Bayesian analyses

Regarding the differences and similarities that we observed using *Visible Vowels*, we will examine whether these impressionistically observed differences are also statistically valid.

### 5.2.1 TTS training vs. TTS

Due to divergence issues, the comparison between the TTS training and TTS corpora was performed using Bayesian linear regression, rather than a

Bayesian mixed-effects model<sup>9</sup>. The formula employed to conduct the Bayesian linear regression is as follows:

$$Formants \sim Register \times Vowel\_types$$

In this analysis, we used normal parameters ( $\mu = 0$ ,  $\sigma = 1$ ), which is a weakly informative prior. The vowel /a/ and the TTS training corpus are set as the intercept (baseline). The model converged well with  $\hat{R} = 1$  for all parameters and MCSE under 0.003, with sufficient effective sample sizes  $> 1000$ ). For the data interpretation, one intuitive way to directly access vowel formant differences between the TTS training corpus and the TTS corpus for each vowel is to look for the parameters showing the interaction between vowel and register (vowel:register). These interaction parameters show the difference in formant values for each vowel from the intercept, namely *a:Training*. The parameter *register[TTS]* signifies the formant value difference for /a/ between the TTS training corpus and the TTS corpus. Since the formant values of, for example, /e/ (other than /a/, which is the intercept) in the TTS training corpus can be calculated from *intercept + vowel[e]*, and /e/ in the TTS corpus can be calculated from *intercept + vowel[e] + register[TTS] + vowel : register[e,TTS]*, the difference in /e/ between these two registers can be found by calculating *register[TTS] + vowel : register[e,TTS]*.

Table 1: Total Register Difference of  $F1$  between Training and TTS  $\alpha = \text{Training}$

Vowel	Total Difference (mean effect)	sd	94% HDI	Interpretation
/a/	0.140	0.119	[-0.085, 0.360]	not significant
/e/	0.030	0.187	[-0.325, 0.382]	not significant
/i/	-0.006	0.126	[-0.233, 0.235]	not significant
/o/	0.128	0.228	[-0.308, 0.543]	not significant
/u/	-0.223	0.131	[-0.468, 0.025]	not significant
/y/	-0.006	0.132	[-0.259, 0.240]	not significant
/ø/	0.053	0.179	[-0.295, 0.382]	not significant
/ɔ/	-0.017	0.181	[-0.351, 0.319]	not significant
/ɛ/	0.107	0.114	[-0.112, 0.313]	not significant

As expected, no parameter shows a significant posterior effect for either  $F1$  or  $F2$ , except for the  $F2$  values of /u/. For the other vowels, when  $\beta$  is negative, the 94% HDI contains positive values, so the difference cannot be interpreted as signifi-

<sup>9</sup>The random effect  $+(1|Speaker)$  is the reason for the divergence in the comparison of the TTS training and TTS corpora. Since there are only four speakers for this comparison, the model does not have enough data to calculate interpersonal variation.

Table 2: Total Register Difference of  $F2$  between Training and TTS  $\alpha = \text{Training}$

Vowel	Total Difference (mean effect)	sd	94% HDI	Interpretation
/a/	0.026	0.108	[-0.178, 0.230]	not significant
/e/	-0.047	0.170	[-0.356, 0.284]	not significant
/i/	0.069	0.115	[-0.143, 0.290]	not significant
/o/	-0.048	0.210	[-0.439, 0.347]	not significant
/u/	-0.330	0.120	[-0.554, -0.106]	significant
/y/	-0.152	0.122	[-0.384, 0.071]	not significant
/ø/	-0.168	0.165	[-0.488, 0.138]	not significant
/ɔ/	-0.164	0.168	[-0.477, 0.151]	not significant
/ɛ/	-0.016	0.107	[-0.222, 0.176]	not significant

cant; similarly, when  $\beta$  is positive, the 94% HDI overlaps with zero. However, for the posterior effect of  $F2$  values of /u/, we can observe that  $\beta = -0.330$  with a 94% HDI of [-0.554, -0.106]. As this interval contains only negative values, this could be interpreted to mean that the difference in /u/ between the two registers is significant.

The overall differences between the TTS training and TTS corpora can be summarised by the following posterior total difference ridge plots:

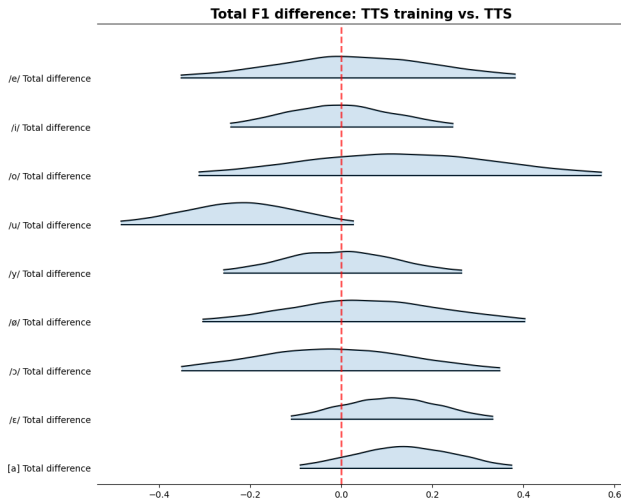


Figure 8: Total difference in  $F1$  between TTS training corpus and TTS

### 5.2.2 Traditional Breton vs. TTS

For this investigation, we used a Bayesian mixed-effects model to adequately capture acoustic variation within traditional speakers ( $n = 10$ ). By accounting for interpersonal variation, we could model the centroid of the traditional speakers. We set normal parameters (weakly informative priors:  $\mu = 0$ ,  $\sigma = 1$ ), consistent with the comparison between the TTS training and TTS corpora. The models (one for  $F1$ , the other for  $F2$ ) fitted well with  $\hat{R} = 1$  for all parameters and MCSE under



Figure 9: Total difference in  $F2$  between TTS training corpus and TTS

0.004, with sufficient sample sizes  $> 1000$ .

The same method was applied as in the comparison between the TTS training and TTS corpora. That is, we will only look for the *total difference* between the traditional speaker corpus and the TTS corpus for each vowel.

Table 3: Total Register Difference of  $F1$  between Traditional and TTS  $\alpha = \text{Traditional}$

Vowel	Total Difference (mean effect)	sd	94% HDI	Interpretation
/a/	0.460	0.131	[0.220, 0.713]	significant
/e/	0.203	0.168	[-0.091, 0.533]	not significant
/i/	0.206	0.136	[-0.040, 0.463]	not significant
/o/	-0.381	0.227	[-0.815, 0.040]	not significant
/u/	0.120	0.144	[-0.146, 0.394]	not significant
/y/	0.172	0.145	[-0.110, 0.436]	not significant
/ø/	-0.617	0.181	[-0.962, -0.286]	significant
/ɔ/	0.016	0.184	[-0.321, 0.365]	not significant
/ɛ/	-0.000	0.129	[-0.243, 0.242]	not significant

Table 4: Total Register Difference of  $F2$  between Traditional and TTS  $\alpha = \text{Traditional}$

Vowel	Total Difference (mean effect)	sd	94% HDI	Interpretation
/a/	0.144	0.099	[-0.044, 0.325]	not significant
/e/	0.257	0.141	[0.001, 0.531]	significant
/i/	-0.004	0.103	[-0.204, 0.182]	not significant
/o/	-0.361	0.195	[-0.731, -0.001]	significant
/u/	-0.230	0.114	[-0.440, -0.015]	significant
/y/	-0.050	0.114	[-0.256, 0.172]	not significant
/ø/	-0.155	0.150	[-0.442, 0.118]	not significant
/ɔ/	-0.082	0.150	[-0.358, 0.202]	not significant
/ɛ/	0.161	0.098	[-0.019, 0.352]	not significant

For  $F2$ , three parameters show significant posterior effects. These are the cases of /e/, /o/, and /u/. Firstly, /e/ shows a positive  $\beta$  ( $= 0.257$ ) with a 94% HDI of [0.001, 0.531]. Certainly, its 94% HDI does not overlap with zero; however, since the interval is quite wide and the lower bound is close to zero, we can say that the difference in /e/ is statistically credible (*i.e.* /e/ in the TTS

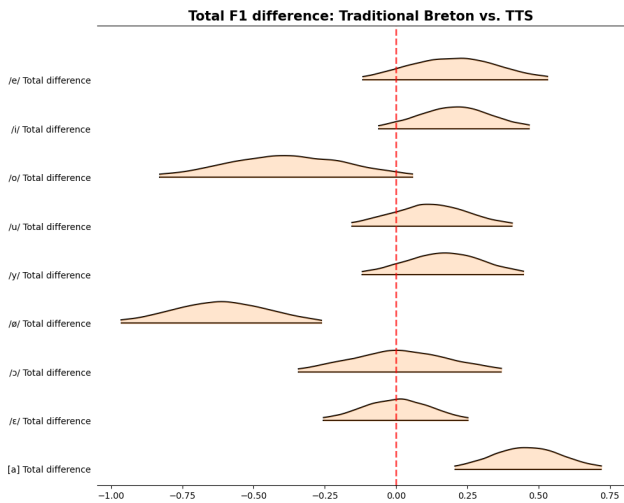


Figure 10: Total difference in  $F1$  between Traditional corpus and TTS

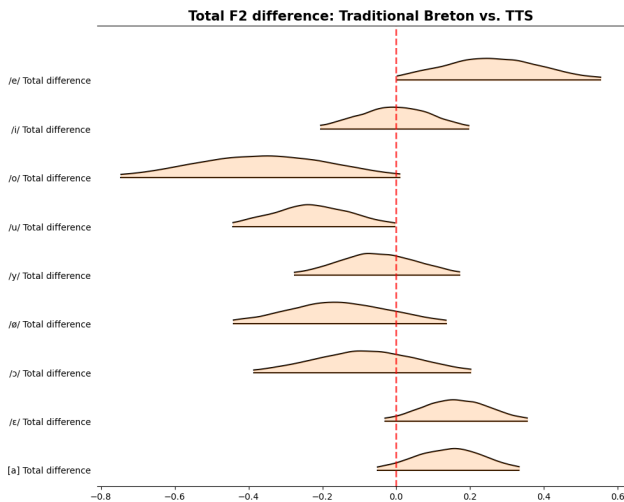


Figure 11: Total difference in  $F2$  between Traditional corpus and TTS

corpus is pronounced with a more fronted tongue position) but is not the most important effect and is marginal. Secondly, /o/ shows the same direction of interpretation as /e/. While /o/ shows a negative  $\beta$  ( $= -0.361$ ) with a 94% HDI of  $[-0.731, -0.001]$  which contains only negative values, since the interval is quite wide and the upper bound is found near zero, the difference concerning /o/ (*i.e.* the TTS corpus pronounces /o/ with a more backed tongue position) is credible but marginal. However, as far as /u/ is concerned, we observe a negative  $\beta$  ( $= -0.230$ ) with a 94% HDI of  $[-0.440, -0.015]$ . As both HDI values are found below zero, we can say that the /u/ difference is credible, since the 94% HDI is positioned relatively far from zero and has a narrow width.

Since our goal is to find obvious differences,  $\beta$  values might be regarded as more important than HDI bounds. Even if larger HDI bounds represent a slightly uncertain effect, a high  $\beta$  value (whether positive or negative) would represent a more significant difference in perception (*the more different, the more difficult to perceive*). From this perspective, we can rank vowel formant differences according to the  $\beta$  of the effective parameters:

Table 5: Total Difference Ranking: Traditional vs. TTS  $\alpha = \text{Traditional}$

Ranking	Vowel Formant	$F$	$\beta$	Interpretation
1	/ø/	$F1$	-0.617	substantially different
2	/a/	$F1$	0.460	markedly different
3	/o/	$F2$	-0.361	pronounced different
4	/e/	$F2$	0.257	shifted
5	/u/	$F2$	-0.230	shifted

## 6 Discussion

Given the high similarity between the TTS training corpus and the TTS corpus, we can say that the engineering fidelity of the Breton TTS well represents its training corpus. However, one significant difference ( $F2$  of /u/) can be found between the TTS training and the TTS corpus. This might be because TTS speech is designed to provide a more stable (less dispersed) vowel range, remaining as faithful as possible to the training corpus. As shown in Figure 7, the TTS shows a more non-overlapping vowel range where possible. This tendency made the TTS pronounce /u/ with lower

$F2$  values (representing a more backed tongue position).

Concerning the difference between the traditional Breton corpus and the TTS corpus, we can say that the TTS satisfies one of the sociolinguistic purposes, which is *intelligibility*. Even if we found some significant effects for some parameters, not all parameters show notable differences, and furthermore, no vowel is different in both  $F1$  and  $F2$ . From this, we could say that sociophonetic representativeness is partially respected, at least for the vowel space. Why are there some effective differences between the traditional Breton and the TTS corpus in terms of vowel space? If we observe carefully the parameters which show a significant difference between the traditional Breton and the TTS corpus, they are (except for /ø/) almost corner or near-corner vowels (/a/, /u/, /e/, and /o/). If we review Figure 7, the vowel range of traditional Breton speakers is quite wide for /e/, /a/, /u/, and /o/, while the TTS training and the TTS corpus have a narrower range. If the TTS corpus' vowel space correctly aligns with the TTS trainers' vowel space, the difference between the traditional and TTS corpus might be just because, in a studio-recording context, trainers are usually asked to produce a clear and distinctive pronunciation, which makes trainers pronounce more *extreme* vowels.

Regarding the difference concerning  $F1$  values of /ø/, one might note a French influence, but our traditional Breton corpus is based on just one specific region (North-eastern Breton), which cannot be absolutely generalised. Also, Bothorel [14, p. 204] noted that in the Argol dialect (closer to North-western dialects),  $F1$  values of /ø/ range widely from 350Hz to 750Hz, which can be closer to our traditional Breton speakers' /ø/ and, at the same time, our TTS trainers' /ø/.

## 7 Conclusion

In our paper, we verified that the Breton TTS ensures engineering fidelity in terms of vowel space, corresponding mostly to the TTS trainers' vowel space. Also, we noted that the Breton TTS largely ensures acoustic intelligibility (at least for the vowel space), which serves an engineering purpose as well as our sociolinguistic purpose. This means

that the Breton TTS could be suitable for commercial or educational purposes. Even though we can note some significant total differences for some vowels (notably for the  $F1$  values of /ø/ and /a/), these do not significantly compromise the sociophonetic representativeness of our main subject, the vowel space.

Regarding future directions, it would be interesting to conduct a further perception study with Breton speakers of various sociolinguistic profiles to evaluate the naturalness of the Breton TTS vowel space. Also, other dialects (especially North-western dialects) could be subjects of comparison with the Breton TTS. Other acoustic cues could be compared with native production, for example, acoustic cues for stress and pitch contour realisations.

## References

- [1] Mélanie Jouitteau and Reun Bideault. “Outils numériques et traitement automatique du breton”. In: *forthcoming*. Ed. by Catherine Schnedecker, Annie Riolland, and Michela Russo. Société Linguistique de Paris, 2023. URL: <https://hal.science/hal-03918268> (visited on 12/21/2023).
- [2] Mélanie Jouitteau. “Guide de survie des langues minorisées à l’heure de l’intelligence artificielle : Appel aux communautés parlantes”. In: *Lapurdum. Euskal ikerketen aldizkaria | Revue d’études basques | Revista de estudios vascos | Basque studies review* 24 (Dec. 14, 2023). ISBN: 9782959007712 Number: 24, pp. 199–217. ISSN: 1273-3830. DOI: 10.4000/127u1. URL: <https://journals.openedition.org/lapurdum/4500?lang=eu> (visited on 02/15/2025).
- [3] Région Bretagne and TMO Régions. *Étude sur les langues de Bretagne : Enquête sociolinguistique*. Synthèse. 2018. URL: <https://www.bretagne.bzh/app/uploads/Étude-sur-les-langues-de-bretagne.pdf>.
- [4] Région Bretagne and TMO Régions. *Enquête sociolinguistique 2024 sur les langues de Bretagne*. Synthèse. 2025.
- [5] Erwan Le Pipec. “Transmission ou transformation ?” In: *La Bretagne Linguistique* 24 (). ISBN: 9791092331561 Number: 24, pp. 5–44. ISSN: 1270-2412. DOI: 10.4000/1b1.4059. URL: <https://journals.openedition.org/1b1/4059> (visited on 03/24/2026).
- [6] David Guennec et al. “BreizhCorpus: A Large Breton Language Speech Corpus and Its Use

- for Text-to-Speech Synthesis”. In: Proc. Odyssey 2022. 2022, pp. 263–270. DOI: 10 . 21437 / Odyssey . 2022 - 37. URL: [https://www.isca-archive.org/odyssey\\_2022/guenec22\\_odyssey.html](https://www.isca-archive.org/odyssey_2022/guenec22_odyssey.html) (visited on 03/31/2026).
- [7] Seongwoo Kang. “Évaluation comparative de l’intonation de la synthèse vocale du breton et de la voix dans la matière pédagogique du breton”. In: *LangaJE* (Oct. 17, 2025). URL: [https://www.academia.edu/143538738/%C3%89valuation1\\_comparative\\_de\\_lintonation\\_de\\_la\\_synth%C3%A8se\\_vocale\\_du\\_breton\\_et\\_de\\_la\\_voix\\_dans\\_la\\_mati%C3%A8re\\_p%C3%A9dagogique\\_du\\_breton\\_LangaJE\\_2025\\_le\\_17\\_octobre\\_2025\\_%C3%A0\\_Lyon](https://www.academia.edu/143538738/%C3%89valuation1_comparative_de_lintonation_de_la_synth%C3%A8se_vocale_du_breton_et_de_la_voix_dans_la_mati%C3%A8re_p%C3%A9dagogique_du_breton_LangaJE_2025_le_17_octobre_2025_%C3%A0_Lyon) (visited on 10/28/2025).
- [8] Seongwoo Kang. “Pitch Accent Distribution in Breton Dialects: from Phonology to Typology”. In: Widame. Maynooth, Nov. 7, 2025.
- [9] Seongwoo Kang. “La prosodie du breton et ses variations dialectales”. In: *La Bretagne Linguistique* (Dec. 6, 2024). URL: [https://www.academia.edu/126369226/La\\_prosodie\\_du\\_breton\\_et\\_ses\\_variations\\_dialectales](https://www.academia.edu/126369226/La_prosodie_du_breton_et_ses_variations_dialectales) (visited on 12/17/2024).
- [10] Mark Liberman and Alan Prince. “On Stress and Linguistic Rhythm”. In: *Linguistic Inquiry* 8.2 (1977), pp. 249–336. ISSN: 0024-3892. URL: <https://www.jstor.org/stable/4177987> (visited on 08/24/2023).
- [11] R. W. Hamming. “Error detecting and error correcting codes”. In: *The Bell System Technical Journal* 29.2 (Apr. 1950), pp. 147–160. ISSN: 0005-8580. DOI: 10 . 1002 / j . 1538 - 7305 . 1950.tb00463.x. URL: <https://ieeexplore.ieee.org/document/6772729> (visited on 07/22/2025).
- [12] Henri Guiter. “Atlas et frontières linguistiques”. In: *Les dialectes romans de France à la lumière des atlas régionaux*. 1973.
- [13] Paolo Mairano et al. *Acoustic distances, Pillai scores and LDA classification scores as metrics of L2 comprehensibility and nativelikeness*. Publication Title: ICPHS2019. 2019. URL: <https://hal.science/hal-03046802> (visited on 01/09/2026).
- [14] André Bothorel. *Etude phonétique et phonologique du breton parlé: à Argol (Finistère-Sud)*. Lille [France] : Spezed [France]: Atelier national, Reproduction des thèses, Université Lille III ; Diffusion Breizh, 1982. 407 pp. ISBN: 978-2-7295-0005-4.
- [15] Olivier Corbin and Jean Léo Léonard. “Esquisse d’analyse de la dispersion allophonique et diatopique des voyelles antérieures hautes et moyennes en breton du Tregor (Ploubezre, Lan-nion)”. In: *L.P.P. Université Paris 3/ Sorbonne Nouvelle CNRS UMR 7018-ILPGA* (2004).
- [16] H. Kennard. “Variation in Breton word stress: new speakers and the influence of French”. In: *Phonology* 38.3 (2021). ISSN: 0952-6757. URL: <https://ora.ox.ac.uk/objects/uuid:e76b4136-3e5a-4bbd-be96-8bbe9844c902> (visited on 02/03/2025).
- [17] James Emil Flege and Ocke-Schwen Bohn. “The Revised Speech Learning Model (SLM-r)”. In: *Second Language Speech Learning: Theoretical and Empirical Progress*. Ed. by Ratrete Wayland. Cambridge: Cambridge University Press, 2021, pp. 3–83. ISBN: 978-1-108-84063-7. DOI: 10 . 1017 / 9781108886901 . 002. URL: <https://www.cambridge.org/core/books/second-language-speech-learning/second-language-speech-learning-revised-speech-learning-model-slmr/7A720FCB65B653B00C766A436908B1A7> (visited on 12/29/2025).
- [18] Tangi Yekel. “Fonologiezh brezhoneg Bear”. Mémoire de Master. Université de Rennes 2, 2025.
- [19] Jean Le Dù. “Le parler breton de la presqu’île de Plougrescant (Côtes-du-Nord)”. PhD thesis. Université de Bretagne occidentale, 1978.
- [20] Paul Boersma and David Weenink. *Praat: doing Phonetics by Computer*. 1992. URL: <https://www.fon.hum.uva.nl/praat/> (visited on 06/20/2022).
- [21] Yannick Jadoul, Bill Thompson, and Bart de Boer. “Introducing Parselmouth: A Python interface to Praat”. In: *Journal of Phonetics* 71 (2018), pp. 1–15. DOI: <https://doi.org/10.1016/j.wocn.2018.07.001>.
- [22] Wilbert Heeringa and Hans Van de Velde. “A New Vowel Normalization for Sociophonetics”. In: Interspeech. 2021. DOI: 10 . 21437 / Interspeech . 2021 - 1846. URL: [https://www.researchgate.net/publication/354221632\\_A\\_New\\_Vowel\\_Normalization\\_for\\_Sociophonetics](https://www.researchgate.net/publication/354221632_A_New_Vowel_Normalization_for_Sociophonetics) (visited on 12/30/2025).
- [23] B. M. Lobanov. “Classification of Russian Vowels Spoken by Different Speakers”. In: *The Journal of the Acoustical Society of America* 49.2 (Feb. 1, 1971), pp. 606–608. ISSN: 0001-4966. DOI: 10 . 1121/1.1912396. URL: <https://doi.org/10.1121/1.1912396> (visited on 03/31/2026).
- [24] Shraavan Vasishth et al. “Bayesian data analysis in the phonetic sciences: A tutorial introduction”. In: *Journal of phonetics* 71 (Nov. 2018), pp. 147–161. ISSN: 0095-4470. DOI: 10 . 1016 / j . wocn . 2018 . 07 . 008. URL: <https://pmc.ncbi.nlm>

nih.gov/articles/PMC6124675/ (visited on 02/20/2026).

- [25] Oriol Abril-Pla et al. “PyMC: a modern, and comprehensive probabilistic programming framework in Python”. In: *PeerJ Computer Science* 9 (Sept. 1, 2023), e1516. ISSN: 2376-5992. DOI: 10.7717/peerj-cs.1516. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10495961/> (visited on 02/20/2026).
- [26] Tomás Capretto et al. *Bambi: A simple interface for fitting Bayesian linear models in Python*. Jan. 11, 2022. DOI: 10.48550/arXiv.2012.10754. arXiv: 2012.10754[stat]. URL: <http://arxiv.org/abs/2012.10754> (visited on 02/20/2026).
- [27] Ravin Kumar et al. “ArviZ a unified library for exploratory analysis of Bayesian models in Python”. In: *Journal of Open Source Software* 4.33 (Jan. 15, 2019), p. 1143. ISSN: 2475-9066. DOI: 10.21105/joss.01143. URL: <https://joss.theoj.org/papers/10.21105/joss.01143> (visited on 03/31/2026).

## 8 Annex: selected sentences for TTS training and TTS vowel analysis

Table 6: Selected sentences for TTS training and TTS vowel analysis

#	Sentence	Code
1	Da bet eur e tigor ar stal?	212_02-001
2	Video: petra a vez graet gant implijidi an hentoù bageal?	212_02-030
3	An dud a chome da labourat evit pennadoù hir pe hiroc’h.	212_03-013
4	Degemer mat er skoazeller grafik!	212_01-002
5	Komprenet eo bet ar gentel-mañ ganeoc’h?	212_02-009
6	Pelec’h emañ o labourat bremañ?	212_02-013
7	Er stal e pren ma zad ar gazetenn bemdez.	212_03-130
8	Aezetoc’h e oa eget na gave din.	212_03-185
9	Torret eo bet ar prenestr gantañ.	212_03-047
10	Ar medisinezed mat.	212_03-186
11	Da amzer a gavez hir.	212_03-035
12	Barbara eo hi.	212_03-042
13	Disheñvel mat eo Chicago diouzh Boston.	212_03-140
14	Pa’z eer d’ar marchad.	212_03-138
15	Meurlarjez e Lannuon.	212_03-020
16	Unan mil uzet.	212_03-197
17	Gwall nebeut a fiziañs ‘meus ennañ.	212_03-070
18	Peseurt micher hoc’h eus?	201_02-052
19	Ha debret hoc’h eus biskoazh soubenn avaloù-douar?	201_05-143
20	Peseurt kelenn a vez kinniget d’ar skolidi e Breizh?	201_01-143
21	Pe daoust ha ne vefe kement-mañ nemet ur c’hliched ouzh-penn?	201_01-019
22	C’hoant hoc’h eus da aotren ar bennadurezh testenian-mañ?	201_05-145
23	Pelec’h emañ ar sal-emwalc’hiñ?	201_01-006
24	Petra eo ho raktres?	201_03-046
25	Gouzout a rit pelec’h emañ ma lunedoù?	201_03-004
26	O terc’hel gouloù-noz en iliz ?	201_01-007
27	Mat eo deoc’h goulenn outañ dont amañ?	201_04-171
28	Da bet eur e oac’h aet da gousket dec’h da noz?	201_03-199
29	Petra a c’hallan ober evidoc’h?	201_01-023
30	Vincent Bollore prest da blantañ reuz?	201_03-009
31	Pelec’h emañ ma roched?	201_01-199
32	Peseurt plas a vo d’ar brezhoneg er c’huzul?	201_02-122
33	Hag asantiñ a rit evit ho pugel lidañ ar baeroniezh keodedel?	201_01-048
34	Marteze e c’hellfes teurel ur sell war ar c’hinnigoù am eus graet evit rubrikennoù kelaouenn Ti-Kêr Plouzane?	201_03-178
35	Kaset hoc’h eus bleunioù din?	201_03-008
36	Ha petra hon boa gwelet neuze?	201_01-049
37	Diwallit da tapout droug stomok dre forzh bezañ marlonk!	205_88
38	Avaloù-douar diblusket bet poazhet en dour da heul.	205_87-1
39	tin , persilh , lore , kignen , hollen ha pebr.	205_60-3

continued

**Table 6 – continued**

---

#	Sentence	Code
40	War un dro gant an houad , e c'heller reiñ rezin diblusket , ha war ar plad en-dro d'an houad , lakaat tammoù krampouezh sall.	205_56-2
41	ar re o deus stourmet leun a spi.	204_02-03
42	hag e ruilhfe 'lec'h ma ruilh karr an ispisour.	204_01-18
43	An dud a zo kousket , evel kent , evel kent , siwazh!	204_04-27
44	O ruilhal, o veuziñ.	204_17-10
45	Klod d'hor stourmerien fuzuilhet ugent vloaz 'zo.	204_17-17
46	O c'hestell gwenn lorc'hus a zo evel skrinoù goullo.	204_17-13

---