

# Using Speech and NLP Resources to build an iCALL platform for a minority language: the story of *An Scéalaí*, the Irish experience to date

**Neasa Ní Chiaráin**

Trinity College, Dublin

Ireland

neasa.nichiarain@tcd.ie

**Oisín Nolan**

Trinity College, Dublin

Ireland

oinolan@tcd.ie

**Madeleine Comtois**

Trinity College, Dublin

Ireland

comtoism@tcd.ie

**Neimhin Robinson Gunning**

Trinity College, Dublin

Ireland

nrobinso@tcd.ie

**Harald Berthelsen**

Trinity College, Dublin

Ireland

berthelh@tcd.ie

**Ailbhe Ní Chasaide**

Trinity College, Dublin

Ireland

anichsid@tcd.ie

## Abstract

This paper describes how emerging linguistic resources and technologies can be used to build a language learning platform for Irish, an endangered language. This platform, *An Scéalaí*, harvests learner corpora – a vital resource both to study the stages of learners' language acquisition and to guide future platform development. A technical description of the platform is provided, including details of how different speech technologies and linguistic resources are fused to provide a holistic learner experience. The active continuous participation of the community, and platform evaluations by learners and teachers, are discussed.

## 1 Introduction

This paper presents our experience in developing an intelligent-Computer-Assisted Language Learning (iCALL) platform for the Irish language, *An Scéalaí* ('The Storyteller'). It promotes the study of Irish, an endangered language, in two distinct ways. Firstly, it deploys linguistic and computational resources to optimise the language learning process. Secondly, it harvests data about how language learners use the platform and stores learners' linguistic compositions, which is crucial

for the study of the Irish acquisition process. The system is complex in that it integrates a number of linguistic and speech resources into a single user-friendly application for learners, while being hosted within a management system that enables high-level guidance by teachers and/or autonomous learning by individuals. A crucial feature is that *An Scéalaí* collects valuable learner data, hitherto unavailable, encompassing both the learners' linguistic output and their engagement with the language tools of the system.

*An Scéalaí* has entailed a cycle of design, implementation, testing, evaluation, redesign, and at the heart of the process has been an extensive collaboration with sectors of the language learning community. As an online system, it has been one of the fortuitous consequences of the global pandemic that an acute appetite for such a resource has resulted in a context which facilitated widespread testing. The present account provides a flavour of the developmental process and discusses the wider potential of this type of platform for many other minority and endangered languages.

*An Scéalaí*, as an iCALL platform, is *intelligent* in that it utilises speech and NLP knowledge and resources in an integrated platform that can opti-

mise the learners' acquisition of the four language skills (writing, listening, reading and speaking) in a holistic way. It is also *intelligent* in capturing the many dimensions of how learners progress in the development of their language skills, providing an intelligent learner corpus (*An Corpus Cliste*). This corpus will guide the future content and platform development.

The platform has involved the integration of different disciplines (linguistics, computational linguistics, engineering sciences) with expertise in the local language and its context. For us, and for many working with minority or endangered languages, getting such an integrated research environment has been very challenging (see more below). As mentioned, a crucial partner in the enterprise has been the active participation of the teaching and learning community. The platform and the experience described here hopefully demonstrates how such interdisciplinary research and development can work alongside a language community, to provide smart learning technologies that will serve future generations of learners and researchers in the field.

## 2 Background

### 2.1 Context: Irish, an Endangered Language

Irish, a Celtic language, is classified as 'definitely endangered' (Moseley, 2012). The communities of native speakers are clustered in small (Gaeltacht) regions, mostly in the West of Ireland. However, as an endangered language, it is unusual in being the first official language. It is a school subject for all, up until school-leaving age (c. 18 years), and hence, there is a large population of learners (c. 700,000 in the Republic of Ireland and unspecified numbers in Northern Ireland) (Ní Chiaráin, 2014). There are also many adult autonomous learners in Ireland and abroad.

There are many challenges for learners of Irish, the most pressing being that most learners do not have ready access to native speakers of the language, or genuine interactions using the language. Teaching resources are often very traditional and often criticized. As in the typical minority language teaching context, the teachers are themselves second language learners. Despite the many challenges, the large numbers of learners presents an opportunity to develop and test systems with large numbers of participants, as evidenced here.

### 2.2 Irish Speech and Language Technology

Despite some flourishing of speech and language resource development for Irish in recent years it remains, in the wider picture, very under-resourced. The lack of speech and language technologies has inevitable consequences in an increasingly digital world. Indeed, this deficit of resources for minority and endangered languages has been described as a digital timebomb (Ní Chasaide et al., 2020), in that it increasingly narrows the domains in which the language can be used, even by native speakers.

*An Scéalaí* is part of a wider initiative, ABAIR, whose mission is the development of linguistic resources, both to document the living language, and to underpin the development of core speech technologies. This initiative is particularly known for the provision of synthetic voices (TTS)<sup>1</sup> for the three main dialects of Irish (note there is no standard spoken variety, a common feature of minority languages). An automatic speech recognition system (ASR)<sup>2</sup> is also at prototype stage. A central concern is the development of the most urgently needed applications, unlike the case of technology in the major languages where development is profit-led. Of particular importance for language maintenance and transmission are sophisticated, interactive educational applications. A further related concern is the provision for those with speech/language or communication difficulties.

### 2.3 Motivation for *An Scéalaí* Design

As described in (Ní Chiaráin and Ní Chasaide, 2019) this platform currently involves the learner in sequential language learning activities. The learner writes some text, a story, and uses the language technologies to self-correct. The text can be listened to, via TTS, providing exposure to native speaker models of the language and enabling proof-listening as a self-correction tool. Spelling and grammar checkers also provide corrective feedback. A link to dictionaries, thesauri and grammar wizards enable the learner to further improve their composition. A facility is also provided for learners to record their own rendition of the story, and to compare it with a native speaker (TTS) rendition. The integration of these tools in the platform is intended to encourage a holistic ap-

<sup>1</sup><https://www.abair.ie>

<sup>2</sup>[https://phoneticsrv3.lcs.tcd.ie/rec/irish\\_asr](https://phoneticsrv3.lcs.tcd.ie/rec/irish_asr)

proach to language learning, where all language skills (writing, listening, reading and speaking) evolve simultaneously and reinforce each other.

Traditionally Irish learning was very text-based, with a focus on syntax, orthography and grammatical accuracy. One of the failures has been the nurturing of the spoken language and, as mentioned above, this has been exacerbated by the fact that most learners do not have access to native speaker models of the language. The inclusion of synthetic voice in a choice of dialect is, in our view, a novel core feature of the platform. Apart from the obvious need to acquire authentic pronunciation skills, it should be noted that the written form of Irish is opaque, in that the link between the sounds and the orthographic forms are complex and typically not grasped by learners or their teachers. By constantly hearing the speech corresponding to their own written text, learners would have much more exposure and more readily grasp the fundamentally phonic structure key to the writing system (Ní Chasaide, 1999).

### 3 Description of the Platform

The platform is rather complex and includes a user-friendly interface where the learner has the benefit of access to feedback based on linguistic and speech resources. This integrated platform is targeting the parallel development of the four language skills (writing, speaking, listening, reading). The system also encompasses software for user and content management, which ensures that the platform is robust and user-friendly and is at all times harvesting learner data. The latter is key to a growing body of learner data, *An Corpas Cliste*, which will be used to study the stages of the acquisition process. This information will enable content development in line with acquisition stages, that can furthermore be personalised to the individual learner.

The platform development was an in-house collaboration where the software was written by our own students. These are pursuing an integrated programme in Computer Science, Linguistics and a Language, where Irish is an option<sup>3</sup>. (Note that this kind of programme provides the researchers with the key interdisciplinary skills and knowledge of the language, a fundamental prerequisite

<sup>3</sup>B.A. in Computer Science, Linguistics and a Language (Irish) <https://www.scss.tcd.ie/undergraduate/computer-science-language/>

for developing sophisticated technologies for minority or endangered languages).

## 3.1 Platform Structure and Technologies Incorporated

*An Scéalaí* not only integrates speech and language resources, but provides a management framework that allows continuous communication between teachers and learners so that personalised guidance can be provided. The various aspects of this system are described here and implemented in a modular system where a set of independent services communicate via a REST API, which functions as one central *An Scéalaí* Node backend.

### 3.1.1 Speech and language technology

#### Text-to-speech

- For **text-to-speech** (TTS) functionality, A REST API is used to access the ABAIR TTS synthesiser (Ní Chasaide et al., 2017), which, when provided with a string of text, returns audio files containing the synthesised speech. The API provides a choice of HMM- and/or DNN-based synthesis in the three main dialects of Irish. Users can select their preferred dialect and speech engine.
- The TTS system also provides timing information about the speech, which is used to produce live text highlighting in sync with audio, to visually connect text and speech.

#### Grammar checker

- An *Gramadóir* (Scannell, 2013) is hosted as a microservice with a REST API that is called directly from the frontend to check text for grammar errors.
- An additional algorithm was added to check for a common spelling error in Irish, to do with vowel agreement within words.
- Further algorithms are being developed to fit with the grammar-checking framework.
- A custom module extending the *Quill*<sup>4</sup> text-editor (see below) was written to enable text highlighting and popup windows over the text. This module is used for displaying *grammar suggestions*, which consist of a text segment specified by *start* and *end* indices, information about the error, which may

<sup>4</sup><https://quilljs.com/>

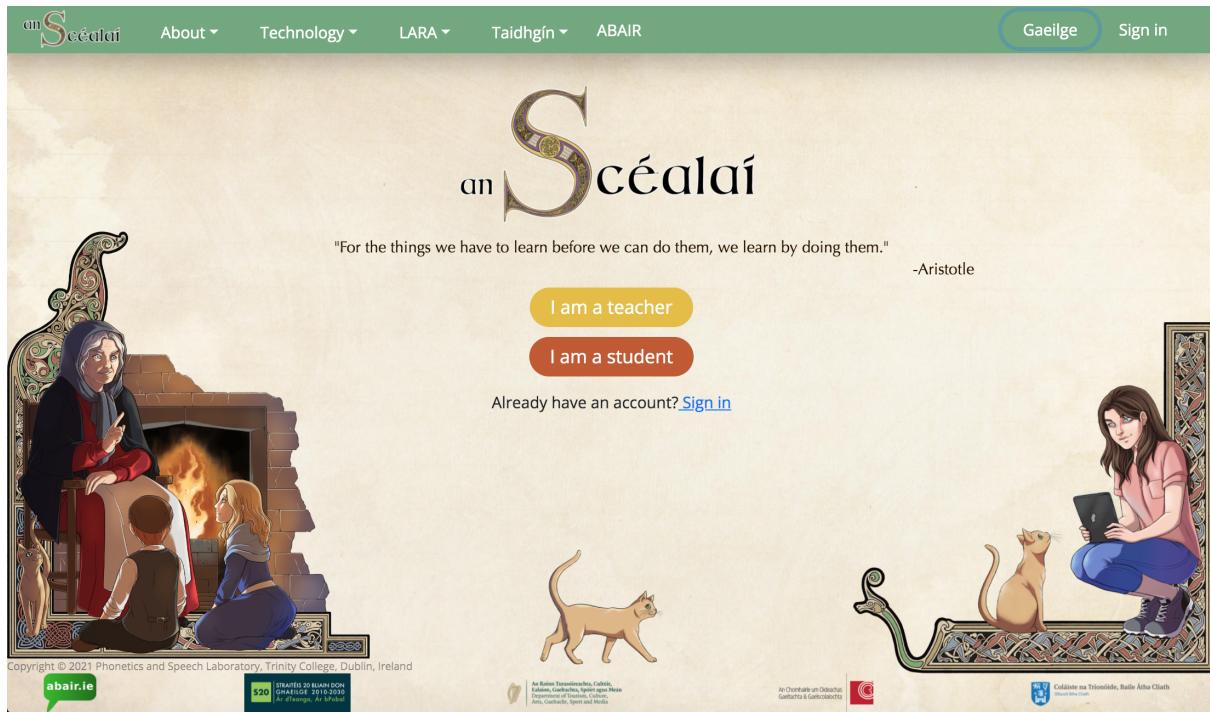


Figure 1: The *An Scéalaí* ('the Storyteller') homepage (English translation). The platform name, the tagline (tarraingíonn scéal scéal eile 'one story begets another') and the imagery on the homepage try to convey that all learners have a story to tell, continuing the well-known Irish storytelling tradition in a modern idiom.

optionally be colour-coded. This encoding of grammar suggestions is designed to be generic, and can be made compatible with a variety of grammar-checking algorithms, rendering a unified and coherent grammar-checker UI on the frontend, while maintaining a modular and extensible set of grammar checking algorithms on the backend.

- Given a grammar error, the highlighting module highlights the specified segment of text in the appropriate colour, and displays further information via popup when the user hovers over a piece of highlighted text.

### Voice recording

- Students can record and listen back to their own voice reading segments of text. Each recording is associated with a piece of text, taken from a snapshot of the story at the time of recording. These recordings can then be archived for future reference, creating a history of voice recordings for a given story over time.
- Each segment of text is also synthesised via TTS, producing a 'gold standard' native

speaker model, to which students can compare their own speech.

### 3.1.2 Managing users & content

The web application was developed using a JavaScript-centric *MEAN* stack, which deploys a MongoDB database, Node.js backend server, Express.js backend framework for API specification, and the Angular framework for the frontend. This choice in tech stack was made for quick prototyping, development, and deployment.

#### User management

- A user may register as either a student or a teacher. They must provide a unique username, password, and e-mail address, and verify these in order to log in.
- Passwords are encrypted using SHA-512 so that passwords are not stored directly on the DB but may easily be validated for authentication.
- User details are stored in a JSON Web Token in the browser's local storage to keep the user authenticated.
- User accounts are assigned a *role* property upon registration (*student*, *teacher*, or *admin*)

and the website presents different views tailored for the different types of user.

- *Classrooms* are effectively sets of students, whose stories a teacher will have access to.

## Content management

- Story data is stored in standard Mongo documents. Additionally, snapshots of the story are saved when students interact with it in certain ways, for example performing a grammar check, or running TTS and listening back. These snapshots are basic elements of the *Corpas Cliste* (see section 3.4).
- The Quill JavaScript library<sup>5</sup> is used as a ready-to-go WYSIWYG editor, basic formatting options in stories. The formatting is encoded to a non-recursize subset of HTML for persistent storage.
- Audio recording is performed using the JavaScript *MediaStream Recording API*<sup>6</sup>, which provides access to user recording devices via the browser. Audio files are stored and retrieved using MongoDB's *GridFS* specification to maintain a more uniform interface to data retrieval.

## 3.2 Student dashboard

- The central student dashboard consists of a text-editor that has been extended to incorporate a grammar-checking tool, text-to-speech synthesis, and a voice recording facility (see Section 3.1.1 for technical detail).
- The student may create and edit multiple texts, or *stories*, using these tools. The stories are associated with their individual accounts and saved on the cloud. They may also be exported to a variety of popular file formats for local storage.
- The text editor provides basic formatting options, producing a familiar writing environment for students.
- Live grammar-checking can be toggled on or off. When switched on, the checker will highlight grammatical errors as they are written. Hovering over a particular error will display information that should help the student

<sup>5</sup><https://quilljs.com/>

<sup>6</sup>[https://developer.mozilla.org/en-US/docs/Web/API/MediaStream\\_Recording\\_API](https://developer.mozilla.org/en-US/docs/Web/API/MediaStream_Recording_API)

resolve it. Students may filter which kinds of errors are flagged using a series of checkboxes below the editor, see Figure 2.



Figure 2: Central student dashboard, with grammar-checking toggled on.

- The story text may be synthesised using the ABAIR TTS system, enabling students to listen back to their story being read aloud. The synthetic utterance may simply be a word, sentence, or a paragraph, allowing students to focus on specific areas of the text.
- A voice recording facility is also provided, so that students may compare their own speech to that of the synthesis. Figure 3 shows the user interface for synthesis and recording.



Figure 3: Students may synthesise their stories, and compare recordings of their own speech.

- Given a unique code, a student may join their teacher's classroom, enabling the teacher to view their stories and provide textual or audio feedback.

## 3.3 Teacher dashboard

- The teacher dashboard is centered around *classrooms*, which are effectively collections of students whose stories the teacher can access and provide feedback on. Students are

notified when they have received either textual or audio (voice-note) feedback from their teacher.

- Each classroom has an associated code with which students can join.

The screenshot shows a list of student usernames on the left and their corresponding error counts on the right. The list includes: seán (4), máire (7), orla (6), clarán (6), cian (3), sinéad (5), and caoimhe (9). At the top of the dashboard, there are buttons for 'Messages' (with a mail icon), 'Stats' (with a bar chart icon), 'Share code' (with a share icon), 'Edit classroom title' (with a pencil icon), and 'Delete' (with a trash bin icon).

Student username	→
seán	→ 4
máire	→ 7
orla	→ 6
clarán	→ 6
cian	→ 3
sinéad	→ 5
caoimhe	→ 9

Figure 4: The classroom dashboard as seen by teacher accounts.

The screenshot shows a student's story titled 'Mo Scéal'. The text of the story is: "Shocraigh Cumann na nGaeilge agus an Cumann Lúthchleas Gael go ndéanfaí athbhreithniú ar na rialacha i mí Eanáir agus chasadair lena chéile arís le gairid le'n é seo a dhéanamh. Cé gur tharraing na rialacha nua conspóid agus go ndearna roinnt daoiné clármsád faoi chuid acu (ach go háirithe an pas líamhne atá imith anois), de réir an méid atá feicthe againn go dtí seo sa chomórtas beimid ag súil leis gur mó an t-airbe ná an t-olc a thioċfaidh na hathruithe." On the right side, there is an 'Information' panel with fields for Title (Mo Scéal), Author (seán), Creation date (08/10/2019, 05:35 PM), and Dialect (connemara). Below this is a 'Feedback' section containing the text 'Go hiontach!' and two buttons: 'Send' and 'Audio message'.

Figure 5: Teachers can directly access their students' stories, where they can leave textual and audio feedback.

- An analytics page provides information on the grammatical errors made by students in the associated classroom. This page also enables teachers to choose which kinds grammar errors are displayed for the students, producing classroom-level configurability for the grammar-checker. This configurability enables teachers to customise their students' experience to fit their lesson plans. In addition to each individual student's statistics, this page also provides an overview of how the class is performing by averaging the kinds and types of errors made by the class as a whole.
- On the messages page, teachers can communicate directly with the students in their classroom in using an interface similar to that of

e-mail systems. They can send either textual or audio messages to individual students or to the entire class. Students who have any questions can then send a message back to their teacher.

### 3.4 The Learner Corpus (*An Corps Cliste*)

- An engagement system is implemented and tracks how *An Scéalaí* is used: each time the grammar checker, TTS, etc. are used, this is logged, along with a timestamp and a copy of the story at that point.
- These logs constitute a rich development history for each story, facilitating analyses of the ways in which TTS, grammar checker etc. are being used, and will allow researchers to examine how they are contributing to student learning, etc.
- Future development for the live grammar checking would be to provide finer time-resolution in grammar checking and correction. Also, in order to allow for more efficient storage, a method to track the difference between two versions of the text, rather than snapshotting the entire texts, will be implemented.

## 4 Community Evaluation > System Enhancement

Experienced Irish language teachers advised on aspects of the initial design of the platform. From prototype stage onwards, extensive consultation and evaluations have been carried out with the community of learners and teachers. The system grows as new/updated technologies come on stream and is being enhanced continuously in response to users' feedback. Groups of users who have contributed evaluations include trainee teachers in Ireland; second level pupils in Ireland; third level students in Ireland; Irish learners in America (part of a Fulbright scheme for Irish teaching in third level institutions in the US); the general public (recruited by word of mouth, as the system is online).

The total number of accounts registered to date is 4,428. The learner corpus now totals 42,542 stories; 5,596,257 words (an average of 131.55 words per story).

### 4.1 Trainee-Teacher Evaluation

We report here on system evaluations, which were carried out on the larger groups ( $n>50$ ). These

involved trainee-teachers at third level and their teachers over the period March – August 2021 (numbers in Table 1). As part of their training, these trainee-teachers are required to spend time among Irish-speaking communities (Gaeltacht), where they take part in an intensive Irish language immersion course. These trainee-teachers are learners of Irish in their own right, and will eventually be teachers of Irish at primary school level. A key element of the course involves a reflective journal, which is periodically reviewed by the teacher and which is used as the basis for an oral interview at the end of the course.

A Gaeltacht-based course was not possible in 2021 due to the pandemic and *An Scéalaí* provided an important core element of an online programme that was provided instead. The design of *An Scéalaí* fortuitously enabled course teachers to keep an overview of work being done in the form of an online reflective journal by learners, and it allowed them to interact with and guide individual learners on an ongoing basis. Note that for this cohort, *An Scéalaí* was being used both as a resource for their own language learning as well as a tool that they will deploy with their own pupils in the future.

We were fortunate in that the course directors engaged enthusiastically and saw the potential of the technology for their students. They collaborated continuously, e.g. facilitating additional workshops so that the system and its workings could be explained to the course teachers. Ongoing communication throughout the duration of the courses ensured that problems arising could be dealt with very promptly and the platform designers were receiving continuous feedback. Additionally, more formal evaluation of both the learner and teacher experience with the platform was elicited through voluntary responses to detailed questionnaires, presented via Google Forms and circulated on the last day of each course.

## 4.2 Learner Questionnaire

Section I elicited:

- learners' previous experience with online learning; the ease of use of *An Scéalaí*; learners' opinions on the usefulness of each of the tools embedded in the platform, using Likert scales and an open comment box.

Section II asked learners:

- in what context they felt *An Scéalaí* would be most useful; what the strengths/weaknesses of the platform are; their suggestions for improving the platform; whether they believed *An Scéalaí* enhanced their learning of Irish; whether it enhanced their confidence as learners; whether or not they'd like to continue using it in future. Open-ended comments were also invited, particularly to elaborate on any negative responses.

## 4.3 Teacher Questionnaire

This elicited teachers' level of experience of teaching online; whether *An Scéalaí* was found to be useful as a management system for the reflective journals; ease of use of the platform from a teachers' perspective; interest in using the platform with other classes in future. An open-ended comment box was included to elicit any feedback (positive/negative) teachers received from students during the course.

## 4.4 From Evaluation to System Enhancement

A total of 1793 learners and 85 teachers registered an account with *An Scéalaí* over the 5-month period. A great deal of information has been gleaned and a glimpse of some of the salient findings are presented here. Responses to some key questions are shown in Tables 1 and 2 below.

Pilot	March	April	May	June	July
(a) No. Accounts	384	293	51	498	603
(b) Respondents	187	222	21	254	494
(c1) Enhanced learning?	89.8% (168)	89.6% (199)	100% (21)	90.2% (229)	91.5% (452)
(c2) Improved confidence?	85% (159)	86% (191)	100% (21)	85.4% (217)	88% (435)
(c3) Use in future?	94.1% (178)	91.9% (204)	100% (21)	87% (221)	90.3% (446)

Table 1: Learner Responses: (a) overall number of accounts registered; (b) number of questionnaires returned; (c) % of positive responses to 3 of the questionnaire items

Responses are overwhelmingly positive from both learners and teachers. This is evidenced by not only the percentage positive responses but also in the open-ended comments, where terms like *réabhlóideach!* ('revolutionary') dominated.

We paid particular attention to any negative feedback and constructive comments emerging in the open-ended comment boxes. Many of these

Pilot	March	April	May	June	July
(a) No. Accounts	18	16	2	21	28
(b) Respondents	8	10	0	7	11
(c1) Useful management system?	100%	100%	-	87.5%	100%
(c2) Use in future?	100%	100%	-	100%	100%
(d) Ease of use?					
V easy:	3	3	-	2	4
Easy:	5	5	-	3	6
Moderate:	0	2	-	2	1
Difficult:	0	0	-	0	0
V difficult:	0	0	-	0	0

Table 2: Teacher Responses (a) number of accounts registered; (b) number of questionnaires returned; (c) % of positive responses to 3 of the questionnaire items; (d) breakdown of responses regarding ease of use on a five-point scale.

were acted on as the pilots were ongoing, during the six month period, so that there was a continuous cyclical process of learner/teacher feedback guiding intensive platform development. Samples of problems highlighted and responses to these problems are included here.

**Text Display:** in the earlier pilots, the most frequent criticism was of the simple plain text editor in the platform (which is the form needed to send data to our TTS servers). Learners wanted a Word-like layout with control over headings, fonts, layout, colour, etc. The solution was to implement a WYSIWYG Editor (see 3.2), which is mirrored by a plain text copy for the synthesis.

**Email verification system:** much time and resources were consumed in the earlier pilots by practical issues, e.g. lost password/lost data requests from learners. An automatic password retrieval system was put in place (see 3.1.2).

**System Robustness:** as the user numbers grew (we had not appreciated that Covid would last so long and that such large numbers would be using this system), the system crashed more frequently. Learners were quick to complain. This is a make-or-break feature to retain student users and, therefore, increasing the robustness is an ongoing priority, requiring continuous bug fixing and extensions to the system. This is a concern for the future, as maintaining this system in the longer term will require ongoing technical support.

**Foregrounding of Oral/Aural skills:** a core objective of this platform is the linkage of the spo-

ken (native) language with the written forms. As mentioned, oral/aural skills have traditionally been neglected in Irish language teaching. *An Scéalaí* should offer a way for parallel development of the four language skills. However, our analysis of the data revealed that the synthetic speech output was relatively little used. While disappointing, on reflection this does not seem particularly surprising: the concept of proof-listening is novel and the platform design relegated this facility to a different tab from the main writing page, making it more likely to be overlooked. To rectify this, the system has been redesigned. In the forthcoming iteration, the option of listening to the spoken output (via TTS) is integrated into the main writing page.

These are relatively large ticket items. Myriads of small fixes were also implemented following feedback, items that would be unlikely to be spotted without user engagement, including: shifting the location of a textbox which obscured the learners' text; and allowing teachers to sort student names alphabetically A-Z, for attendance keeping.

## 5 Conclusions

Successful language transmission is key to language maintenance and revitalisation. In the Irish context, where the population at large receives Irish language instruction, there is the potential to make Irish a vibrant second language beyond the native speaker (or Gaeltacht) population. Computational approaches and the incorporation of linguistic and technology resources can turn the digital timebomb into a major stimulus to language pedagogy and indirectly to language maintenance. Besides its immediate use as a pedagogical platform, *An Scéalaí* will enable documentation and analysis of the stages of acquisition and it will stimulate future development and increasingly effective, technology-based interventions, where our linguistic knowledge is brought to bear.

Our experience has been that the language community is central to all aspects of the wider ABAIR initiative — in this case, the engagement of the educational sector — not a passive recipient but a vital partner in the enterprise, involved in every aspect from design to evaluation and dissemination.

The current platform is a work in progress. It integrates speech and language knowledge as well as core language technologies to provide a holistic learning environment. As further resources, such as ASR, come on stream, the aim is to expand the

capacity of the platform in ways that will further enrich the learners' experience.

### 5.1 Future Directions

The platform as currently configured is potentially language independent (see below). A current focus is to increase the Irish language content. Towards this, we are actively investigating the use of avatars as a way of enhancing the delivery of the spoken output, including feedback on errors, etc. We are also developing independent grammar checking modules to cater for items not detected by the grammar checker currently in use, such as the genitive case marking. The system is also being linked to an interactive chatbot (e.g. to practice irregular verbs); a story starter (to kickstart the writing process); dictogloss (a text reconstruction exercise). A parallel development we have been involved with, the Learning and Reading Assistant (LARA) (Zuckerman et al., 2021), is currently integrated with *An Scéalaí*. The intention is to expand the content offered with a view to encouraging learners to read for pleasure.

Our future wishlist will include a redesign of the system for the mobile phone/tablet, as our analytics data shows that many learners are logging in using mobile phones and tablets. It is frequently pointed out that mobile phones/tablets are much more useful in a minority/endangered language context, where users may not have laptops.

### 5.2 *An Scéalaí*: a Model for Other Languages?

We increasingly see ourselves as part of a global movement where minority and endangered languages share common cause. To this end, *An Scéalaí* is developed as an open source platform, available on GitHub<sup>7</sup>. At its core it has a modular design. Once a given resource (such as TTS, dictionary, etc.) is available in a language, it should in principle be possible to clone *An Scéalaí* for a different endangered language, by slotting the resource in to the core framework. It would be particularly rewarding to find that we can share our experience and resources with other endangered language groups who strive to document, maintain and revive their linguistic heritage.

<sup>7</sup><https://github.com/OisinNolan/An-Scealaí>

## 6 Acknowledgements

Funding from the Department of Tourism, Culture, Arts, Gaeltacht, Sports and Media, Government of Ireland (*ABAIR*) and An Chomhairle um Oideachas Gaeltachta agus Gaelscolaíochta (*An Corpas Ciste*) is gratefully acknowledged. We would also like to thank Michelle de Bhailís and the community at Coláiste Ghaoth Dobhair for their enthusiastic involvement in the project to date.

## References

- Christopher Moseley. 2012. *The UNESCO atlas of the world's languages in danger*, 3<sup>rd</sup> edition. World Oral Literature Project.
- Ailbhe Ní Chasaide. 1999. Irish. In *Handbook of the International Phonetic Association: a guide to the use of the international phonetic alphabet*, pages 111–116. Cambridge University Press, Cambridge.
- Ailbhe Ní Chasaide, Neasa Ní Chiaráin, Christoph Wendler, Harald Berthelsen, Andy Murphy, and Christer Gobl. 2017. The abair initiative: Bringing spoken irish into the digital space. In *INTERSPEECH*, pages 2113–2117.
- Ailbhe Ní Chasaide, Neasa Ní Chiaráin, Harald Berthelsen, Christoph Wendler, Andrew Murphy, Emily Barnes, and Christer Gobl. 2020. Can we defuse the digital timebomb? linguistics, speech technology and the irish language community. In *Proceedings of International Conference on Language Technologies for ALL (LT4All): Enabling Linguistic Diversity and Multilingualism Worldwide*, pages 177 – 181, UNESCO, Paris, France.
- Neasa Ní Chiaráin. 2014. *Text-to-Speech Synthesis in Computer-Assisted Language Learning for Irish: Development and Evaluation (Unpublished doctoral dissertation)*. Ph.D. thesis, The School of Linguistic, Speech and Communication Sciences, Trinity College Dublin.
- Neasa Ní Chiaráin and Ailbhe Ní Chasaide. 2019. An icall approach to morphophonemic training for irish using speech technology. In *CALL and complexity - short papers from EUROCALL 2019*, pages 314 – 320.
- Kevin Scannell. 2013. *An Gramadóir: an open source grammar checking engine, Version 0.70*. <https://cadhan.com/gramadoir/>.
- Ghil'ad Zuckerman, Sigrún Ó Vigfússon, Manny Rayner, Neasa Ní Chiaráin, Nedelina Ivanova, Hanieh Habibi, and Branislav Bédi. 2021. LARA in the service of revitalistics and documentary linguistics: Community engagement and endangered

languages. In *Proceedings of the Workshop on Computational Methods for Endangered Languages*, volume 1, pages 13–23.