Celtic CALL: Strengthening the Vital Role of Education for Language Transmission

**Neasa Ní Chiaráin, Madeleine Comtois, Oisín Nolan, Neimhin Robinson-Gunning,
John Sloan, Harald Berthelsen, Ailbhe Ní Chasaide**

Trinity College Dublin
The Phonetics and Speech Laboratory, School of Linguistic, Speech and Communication Sciences

{neasa.nichiarain, comtoism, oinolan, nrobinso, sloanjo, berthelh, anichsid}@tcd.ie

Abstract

In this paper, we present the Irish language learning platform, *An Scéalaí*, an intelligent Computer-Assisted Language Learning (iCALL) system which incorporates speech and language technologies in ways that promote the holistic development of the language skills - writing, listening, reading, and speaking. The technologies offer the advantage of extensive feedback in spoken and written form, enabling learners to improve their production. The system works equally as a classroom-based tool and as a standalone platform for the autonomous learner. Given the key role of education for the transmission of all the Celtic languages, it is vital that digital technologies be harnessed to maximise the effectiveness of language teaching/learning. *An Scéalaí* has been used by large numbers of learners and teachers and has received very positive feedback. It is built as a modular system which allows existing and newly emerging technologies to be readily integrated, even if those technologies are still in development phase. The architecture is largely language-independent, and as an open-source system, it is hoped that it can be usefully deployed in other Celtic languages.

Keywords: Irish, intelligent-Computer-Assisted Language Learning (iCALL), Modular Design

1. Introduction

This paper introduces an iCALL platform for the teaching/learning of Irish - *An Scéalaí*, ('the Storyteller'), available at abair.ie/scealai. It is built as a modular system which integrates speech and language technologies as they emerge and strives to enable parallel development of all language skills, including speaking, writing, listening and reading. Section 4 below gives a detailed overview of the system architecture, describing its modular nature and discussing how it can, in principle, be used as a language-independent platform that could be deployed by other Celtic language communities. Sections 2 and 3 first situate this iCALL platform in the current Irish educational context and then in the wider context of speech and language technology resource development for endangered and minority languages.

2. The Irish language educational context

In February 1922 the Provisional Government of the newly established State placed the Irish language at the centre of their vision for education in the Free State. From the beginning of the independent State the Irish public supported the 'expectation that the Gaelicisation of Ireland... would be achieved through its education system' (Hyland and Milne, 1992).
Despite this, the outcomes of Irish language education in the majority of schools (English-medium) are poor. Over the last two decades, the situation has continued to worsen and pupils' learning outcomes in Irish are still of concern (Department of Education, 2022).

There are many challenges. The Irish speaking community is quite small. A recent report commissioned by Glór na nGael, estimated that there were c. 7,000 Irish speaking families, including some 16,000 children in the whole of Ireland, with a quarter located in Gaeltacht areas (Seoighe et al., 2021). Effectively the teacher carries enormous responsibilities for the transmission and promotion of the language, particularly at primary level. Most teachers are themselves second language learners and there are issues concerning their own motivation levels and proficiency (Dunne, 2019). Learner engagement is critically dependent on the classroom teacher (Devitt et al., 2018) but many teachers feel that they are poorly supported in this important aspect of their mission (Dunne, 2019).

For most learners the classroom presents the only opportunity to connect with the language (Ó Murchú, 2016) and the majority never have an opportunity to converse with a native speaker. Thus, pupils have insufficient access to native speaker models of the language. An aspiration in the 2022 Chief Inspector's Report is to "develop pupils' academic, cognitive and social language to enable them to use the language more independently, confidently and creatively". This report also recommends that "schools should make further use of school self-evaluation and assessment processes to develop pupils' literacy and communication skills in Irish to support their accurate use of the language" (Department of Education, 2022).

Related to the above, the teaching of the spoken language and pronunciation is frequently seen as a particular failure in Irish language teaching. Most learners

have a poor grasp of the sound system of the language and little sense of how the sounds relate to the writing system (the phonics of Irish).

When it comes to reading, there is often a complete focus on a single textbook for the entire year and reading for pleasure is typically not considered. The materials used for teaching often tend to compare badly with the attractive, interactive materials available in other subject areas.

Given the short time allotted to Irish lessons, (in English-medium schools) and the high pupil-teacher ratio, there is limited opportunity for the individual to engage with the teacher and to get personalised feedback on progress.

3. The Role of Speech and Language Technology

Despite the dire situation of current Irish language education, it remains true that the population at large has a positive attitude to the language. This is reflected partly in the rapid growth of Irish medium schools, which has grown to 8.1% of the total number of schools (Gaeloideachas, 2022). Furthermore, there is a very positive attitude to the use of technology in the classroom.

The *An Scéalaí* platform described here is part of a broader initiative (ABAIR, 2022), involving the development of speech and language resources. Core technologies developed to date include synthetic (male and/or female) voices for the three main dialects and a first speech recognition system is now available (ÉIST, 2022).

A core part of ABAIR's mission is to serve the needs of the language community and consequently, in parallel with core technology development, applications are being built for the public, for those with disabilities and for Irish language education. Initial exploration in this latter area has included proof-of-concept development of interactive language learning games, such as *Taidhgín*, an animated chatbot and *Digichaint*, an interactive adventure game. These proved to be popular with school-going pupils and showed an appetite for this approach (Ní Chiaráin and Ní Chasaide, 2020). Building on these prototypes, we are now developing a comprehensive platform, *An Scéalaí*, that exploits all the technologies currently available in an integrated platform, that is user-friendly for both learners and teachers. This learning platform is also seen as a research tool which will harvest learner and teacher data, leading to iterative longer-term development of intelligent-CALL (iCALL) for the Irish language.

In recent years, the gap between the performance of high and low resource speech and language technologies has widened. This is due, primarily, to the vast amounts of data required by the deep learning models which have generated the improvements in high resource languages (Lugosch et al., 2019). The same levels of data are typically not currently available for low

resource languages. Additionally, while high resource languages have a large pool of expertise to call on, low resource languages may have few, or even none in certain areas. This has led to a significant relative deficit in the available resources, which has been described as a *'digital timebomb'* for those languages that cannot keep pace (Ní Chasaide et al., 2020).

For developers of CALL applications for low-resource languages, the availability and quality of speech and language technologies is often the deciding factor in the functionality which can be presented to end users. If a language has a great speech recognition engine, but lacks synthesis, a CALL application using the available resources will have to focus on the affordances provided by the recognition, e.g. pronunciation training. Conversely, a language with strong synthesis but not recognition will be inclined to favour listening exercises. The design of the application will necessarily follow the resources at hand.

It is not possible to predict whether the quality of speech and language technologies for low resource languages will catch up to the standard being set by the high resource ones. This uncertainty means development of CALL applications for low resource languages is governed, in large part, by what is usable now. This leads to two pertinent issues. First, uncertainty needs to be built in to CALL applications for low resource languages. They need to be ready to incorporate new or improved functionality as soon as it arrives. If, for example, an application was built purely for a language with no speech technologies, but then a good recognition engine became available, the application should be structured in a way that this new functionality could easily be slotted in. Therefore, CALL applications for low resource languages need to be extremely adaptable. Second, these platforms should be constructed to be as language independent as possible. The elements of a CALL platform for one language should be made easily portable for another. This should be particularly the case for closely related languages, e.g. the Celtic family.

In this paper, we introduce *An Scéalaí*, an open source CALL application for Irish which has been designed to be language independent and highly adaptable to future changes in speech and language technologies. It aims to serve both as a template of a successful, practical CALL application for a low-resource language, and also as a codebase for developers to clone and slot in their own resources.

4. The *An Scéalaí* Platform

An Scéalaí (abair.ie/scealai/) is a web application where learners can write stories, listen to a synthetic voice read their story in any one of the three main dialects, record their own voice, consult a dictionary, get feedback from teachers, and receive automated grammatical feedback on common errors. As shown in Figure 2, all this functionality is available by clicking on the re-
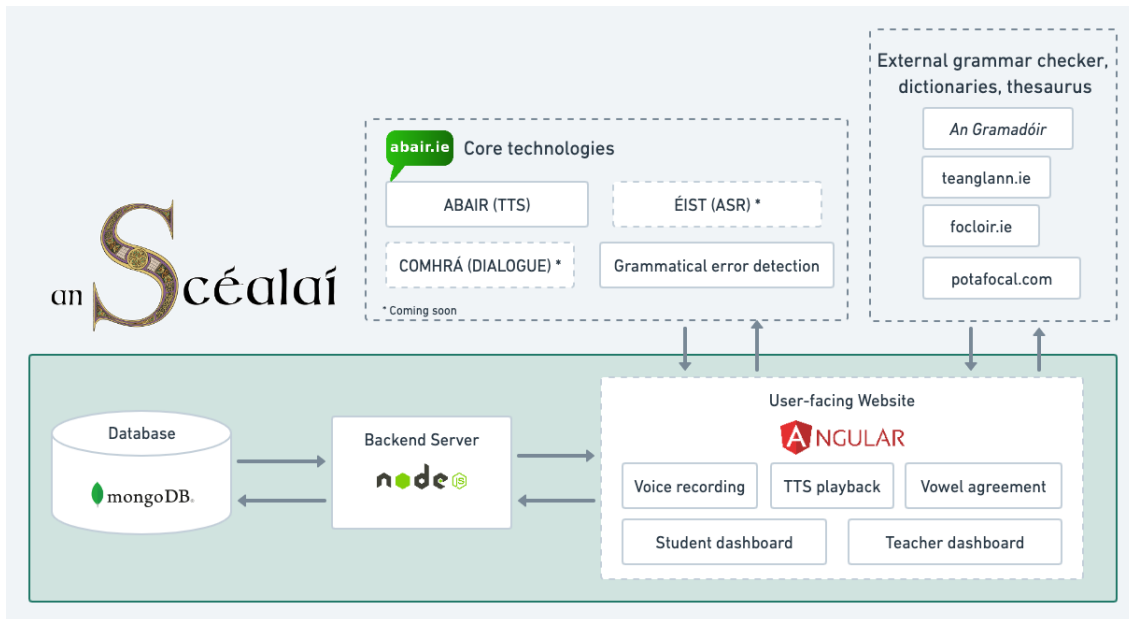
Figure 1: System architecture: an overview of how individual components are combined in *An Scéalaí*.
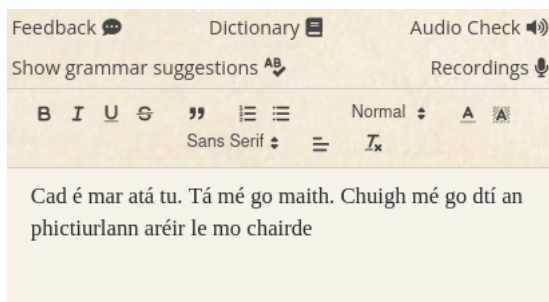

Figure 2: Learner interface for *An Scéalaí* (English version).

lated buttons positioned above the learner's story. Since its inception in 2019, *An Scéalaí* has seen over 4,000 learners write over 40,000 stories, totalling over 5 million words.

The popularity of *An Scéalaí* is due, in part, to its appeal to both sides of the educational divide. Learners are attracted to the platform because it provides spoken synthesis of their own sentences with instantaneous grammatical feedback. Irish, as a Celtic language, contains complex pronunciation and grammatical rules which are quite different from most learners' native language (typically English). Irish has a complex sound system and an opaque writing system. For the learner of Irish (and most teachers), the link between the sounds and the written forms is generally not appreciated. This can make it difficult for the learner to determine correct pronunciation from written text alone, and certain 'basic' grammatical errors persist in many learners' production even at intermediate and advanced levels. *An Scéalaí* provides these learners with

opportunities to privately self-correct and improve their production. These features answer to the aspiration set out by the 2022 Chief Inspector's Report to enable pupils "to use the language more independently, confidently and creatively", using self-evaluation to develop literacy, communication skills and a more accurate use of the language (Department of Education, 2022)

An Scéalaí has proven popular with pupils but also with their teachers. Self-correction facilities reduces the teachers' workload, as it guides learners towards native-like pronunciation and grammatical forms, even without their direct intervention, and the drafts submitted to them are of a much higher standard. As most Irish teachers are not, themselves, native speakers, many report using the platform to check their own notes/feedback before sending it on to learners or parents.

4.1. Modular Approach

An Scéalaí has been designed with modularity at the heart of its design to allow flexibility for changing technologies, user requirements and personnel (see (Figure 1) for an overview of how the various components are combined). It is built using the Angular web framework (https://angular.io/), which utilizes *modules* and *components* to separate functionality into discrete building blocks. These can be developed and tested independently, then easily inserted into the main application. This structure has allowed a diverse range of people to actively contribute to the project, e.g. undergraduate and graduate students, web developers, software engineers, etc. It also has enabled the rapid incorporation of teacher and learner feedback into the platform. Each of these individual modules and their functionality is described below.

4.2. Text-to-Speech Synthesis

A REST API call is made to generate the speech synthesis from the sentences written by the learner. These are available in the three main dialects of Irish through the ABAIR TTS engine (Ní Chasaide et al., 2017). On the learner interface, they appear as buttons beneath the main story (Figure 3). When clicked, they play an audio file. This functionality is designed to guide the learner towards native-like pronunciation in Irish. Common pronunciation problems emerge from an over-reliance on the mapping of the English sound system onto Irish orthography. Most Irish learners do not have ready access to native speaker pronunciation, especially on examples of their own spontaneous output.
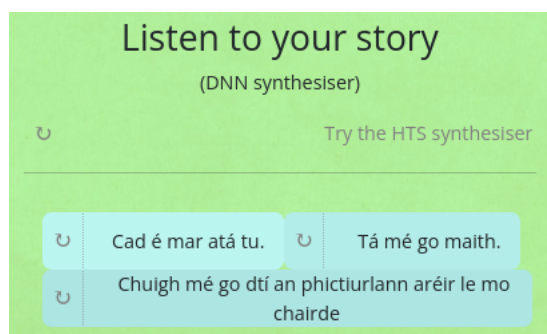

Figure 3: Buttons play synthesised audio for each sentence (English version).

4.3. Recording Audio

To increase the benefit provided to the learner through listening to the synthetic voice, there is also functionality included to allow recording and listening to their own voice (see Figure 4). They can then compare this recording to the synthesised audio. This is achieved by using the browser's MediaStream Recording API.
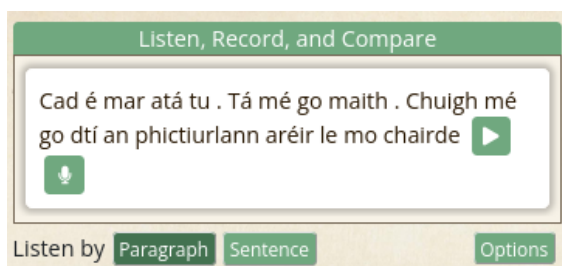

Figure 4: Options for recording one's own voice, listening back and comparing to the (synthetic) native-like version (English version).

4.4. Dictionary

The teanglann Irish dictionary (www.teanglann.ie) is included as a HTML *i-frame*. It appears beneath the

story on the main interface and allows learners to access this resource without needing to navigate away from the page (Figure 5).
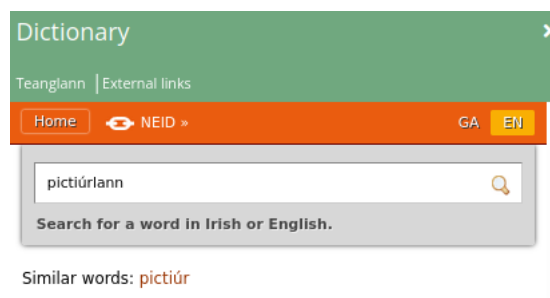

Figure 5: The *teanglann.ie* dictionary is available to search while working on a composition (English version).
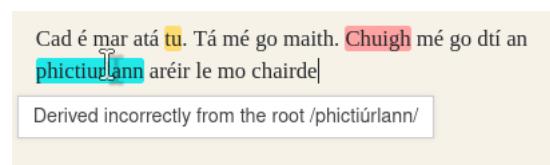
4.5. Automated Grammar Correction

Figure 6: Errors highlighted with colour coding. Suggestions available on hover (English version).

Learners have access to the *An Gramadóir* (Scannell, 2013) grammar checker. Clicking the *show grammar suggestions* button sends a REST API call to the resource, which returns the location, error type, and suggestions as to the nature of the error and how it may be resolved (see Figure 6). Additional algorithms are being added to help with common spelling errors, and more are currently under development to deal with more complex grammatical structures not covered by *An Gramadóir*. These errors are then displayed to the learner by highlighting words (colour-coded by error-typev), with suggestions available if hovered by the cursor (see Figure 6).

4.6. Teacher Feedback

Teachers can create classrooms and assign learners to their class. Here, they are able to view the learners' stories and send feedback. This is then available to the learner through the feedback button on the main interface (Figure 7).

5. Structure for Modification and Developments

The *An Scéalaí* platform is structured so that individual speech and language technologies can easily be inserted and removed from the main story interface. In
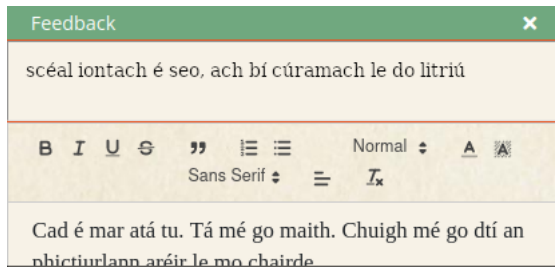
Figure 7: Teacher feedback on the story displayed to the learner in upper window (English version).

each example in the previous section, the relevant technology was displayed to the learner beneath their story. There is no interference from the individual technologies on each other, or the core function. Rather, they serve to enhance the learning experience and outcomes. One notable absence in the technologies available on *An Scéalaí* so far is speech recognition. Until recently, the word error rate of the ABAIR-ÉIST speech recognition (ASR) system was too high to enable inclusion. However, the rapid advances in our current system mean we expect it to be of a sufficiently high standard to be included as a module in the main *An Scéalaí* interface (see (Lonergan et al., 2022) for more detail). This will open up many new opportunities for learning activities. For example, it will make the *An Scéalaí* platform accessible for those who do not (yet) have typing skills / literacy difficulties. It will also open up the field of Computer-Assisted Pronunciation Training (CAPT) comparing learner utterances with the native speaker models. Most of all, it will form the crux of dialogue capacity, enabling the learner to have spoken interactions with virtual native speakers. This will help alleviate the currently limited opportunities learners have to engage in spoken communication.

Due to the modular design, this addition can largely be modelled on those already added. A learner can click a button, and beneath their story a microphone button for recording their voice will appear. After recording, a REST API call to the recognition service will be made, and the text delivered in response can be displayed. This then leads on to possibilities for additional functionality related to pronunciation, particularly when the resulting text does not match the target. Whatever form this will take, it can be developed independently and added to the main interface in exactly the same way.

6. Conclusion

An Scéalaí has now been used by large numbers and formal evaluation is currently being processed. Although it will take time to compile the results, it is clear that the response is overwhelmingly positive from both learners and teachers (some preliminary results are presented in (Ní Chiaráin et al., 2022)).

The beauty of this platform is that it has a simple, modular architecture, which means that as the technologies

evolve, they can easily be incorporated to enhance its scope. As an open source platform, built with replicability in mind, we hope that it can be deployed by other Celtic languages, regardless of the level of currently available speech and language technology resources. Our experience tells us that even while resources are at a very rudimentary stage, it can have a big impact on the learning process. The key factor is not necessarily the technology but that its use is guided by the pedagogical aims and the platform and technologies are developed in partnership with the pedagogical and linguistic experts.

Our experience has also shown that even a very embryonic prototype tends to generate interest and a demand for more development, with a snowball effect. This draws in more teachers, learners, content developers, and so on.

We would like to think that *An Scéalaí* matches the aspirations of the Department of Education cited above. It marries well with the current curriculum and reflects our growing capacity to use digital technology to support teaching and learning. Ultimately, we are hopeful that *An Scéalaí* will contribute to the effective transmission of the language, whether in Irish- or English-medium schools, or in the context of the autonomous learner.

7. Acknowledgements

8. Bibliographical References

ABAIR, (2022). *ABAIR: An Sintéiseoir Gaeilge - The Irish Language Synthesiser ABAIR*. www.abair.ie. As of 14 April 2022.

Department of Education, (2022). *Chief Inspector's Report, September 2016 – December 2020*.

Devitt, A., Condon, J., Dalton, G., O'Connell, J., and Ní Dhuinn, M. (2018). An maith leat an Ghaeilge? An analysis of variation in primary pupil attitudes to Irish in the Growing up in Ireland study. In *International Journal of Bilingual Education and Bilingualism*, volume 21, pages 105–117.

Dunne, C. M. (2019). Primary teachers' experiences in preparing to teach Irish: views on promoting the language and language proficiency. *Studies in Self-Access Learning*, 10(1):21–43.

ÉIST, (2022). *ÉIST: Aithint Chainte na Gaeilge: The Irish Language Recogniser ÉIST*. https://phoneticsrv3.lcs.tcd.ie/rec/irish_asr. As of 14 April 2022.

Gaeloideachas, (2022). *Statistics.* `https://gael oideachas.ie/i-am-a-researcher/sta tistics/`. As of 14 April 2022.

Hyland, Á. and Milne, K., (1992). *Irish Educational Documents, Volume II, Dublin: C.I.C.E.*

Lonergan, L., Qian, M., Berthelsen, H., Murphy, A., Wendler, C., Ní Chiaráin, N., Gobl, C., and Ní Chasaide, A. (2022). Automatic Speech Recognition for Irish: the ABAIR-ÉIST system. In *The 4th Celtic Language Technology Workshop, LREC 2022.*

Lugosch, L., Ravanelli, M., Ignoto, P., Tomar, V. S., and Bengio, Y. (2019). Speech model pre-training for end-to-end spoken language understanding. *arXiv preprint arXiv:1904.03670.*

Ní Chasaide, A., Ní Chiaráin, N., Wendler, C., Berthelsen, H., Murphy, A., and Gobl, C. (2017). The ABAIR Initiative: Bringing spoken Irish into the digital space. In *INTERSPEECH*, pages 2113–2117.

Ní Chasaide, A., Ní Chiaráin, N., Berthelsen, H., Wendler, C., Murphy, A., Barnes, E., and Gobl, C. (2020). Can we defuse the digital timebomb? linguistics, speech technology and the irish language community.

Ní Chiaráin, N., Nolan, O., Comtois, M., Robinson Gunning, N., Berthelsen, H., and Ní Chasaide, A. (2022). Using Speech and NLP resources to build an iCALL platform for a minority language: the story of *An Scéalaí*, the Irish experience to date. In *The Fifth Workshop on the Use of Computational Methods in the Study of Endangered Languages, ACL 2022.*

Ní Chiaráin, N. and Ní Chasaide, A. (2020). The potential of text-to-speech synthesis in computer-assisted language learning: A minority language perspective. In Alberto Andujar, editor, *Recent Tools for Computer- and Mobile-Assisted Foreign Language Learning*, chapter 7, pages 149–169. IGI Global, Hershey, PA.

Scannell, K., (2013). *An Gramadóir.* `https://ca dhan.com/gramadoir/`. As of 14 April 2022.

Seoighe, S., Smith-Christmas, C., and Ó hIfearnáin, T., (2021). *Líon agus Lonnaíocht na dTeaghlach atá ag tógáil clainne le Gaeilge lasmuigh den Ghaeltacht.* `https://www.glornangael.ie/wp-cont ent/uploads/2022/04/Lion-agus-Lonn aiocht-na-dTeaghlach-a-labhraionn- Gaeilge-19-Samhain-2021.pdf/`. As of 14 April 2022.

Ó Murchú, H. (2016). *Irish: The Irish language in education in the Republic of Ireland.* The Netherlands: Mercator European Research Centre on Multilingualism and Language Learning, Regional Dossiers Series, Cambridge.