Mol an Óige: a phonological awareness and early literacy platform for Irish

Ailbhe Ní Chasaide1, Neasa Ní Chiaráin1, Rian Errity1, Oskar Mroz1, Ornait Ní hAonghusa1, Sibéal Ní Chasaide1, Anna Giovannini1, Emily Barnes2

1Phonetics and Speech Laboratory, School of Linguistic, Speech and Communication Sciences, Trinity College Dublin, Ireland
2School of Education, Trinity College Dublin, Ireland

anichsid@tcd.ie, ebarnes@tcd.ie, nichiarn@tcd.ie

Abstract
The Mol an Óige platform for the young learner of Irish, targets phonological awareness and early literacy development, areas of particular difficulty for learners. Firstly, the phonological distinctions of Irish, very different from English, are trained through an immersive environment, where characters bring minimal pairs to life, with songs and stories to aid consolidation, and with quizzes to establish whether the contrasts are acquired. Secondly, the sound contrasts are related to their orthographic representations, and games provide training on the allowed combinations of orthographic letters. TTS is used initially in games involving encoding and decoding of syllables, words and phrases, and eventually to accompany the composition and reading of longer texts. For the future, the provision of children’s TTS and ASR systems is a priority. The central importance of linguistic and pedagogical expertise highlights the need for an interdisciplinary approach to development.

Index Terms: phonological awareness, Irish educational platform, TTS, feedback

1. Introduction
The pronunciation of Irish is widely seen as the least satisfactorily taught aspect of Irish, and many struggle with the (on the face of it) impenetrable writing system. This paper presents a platform that is being built for young learners, to equip them with an awareness of the sound contrasts of the language and with the basic phonic principles of the writing system. The platform addresses a fundamental gap in provision for Irish language learners. And although designed for the young learner, the content is widely needed, and the older learner will also find it useful. The platform can be used on a tablet, phone, classroom whiteboard or computer. The emphasis is on making it game-based and attractive to children. As well as serving as a classroom-based learning tool, it is hoped that children (and others) will use it at home, on their own and with their parents.

This work is part of the wider programme of the ABAIR initiative (see www.abair.ie) which is developing speech technologies and linguistic resources for Irish and which in parallel exploits both technology and linguistic resources to build applications, geared at language learning and at the provision of accessibility systems for those with disabilities.

The platform uses illustrated stories and songs to train the learner’s ear, while games and puzzles reinforce and test learner’s acquisition in a series of steps. The first two steps use pre-recorded materials, mainly stories, songs and prompts to the games. Later steps entail more creative encoding and decoding, and deploy the ABAIR synthetic voices (available for the three main dialects at www.abair.ie). These voices provide spoken prompts and feedback for the learner’s writing attempts as well as the audio tracks to accompany the learner’s reading efforts.

This open-ended availability of spoken output of all written and reading materials is especially valuable, providing ongoing exposure to native-speaker models of the language – something hitherto lacking for most learners of Irish. This ensures ongoing reinforcement of native pronunciation, supplementing and reinforcing the explicit learning of phonological contrasts and the phonic correspondence of sounds to letters. The linguistic components developed as part of the TTS systems for the different dialects are also useful. These includes the dialect-sensitive letter-to-sound conversion rules and dialect lexicons catering for exceptions to the general phonic principles of Irish.

2. The Challenges
Irish is an endangered language, spoken as a community language in small Gaeltacht areas, mostly situated along the Western seaboard of Ireland. Other than very young preschool infants, Irish speakers are bilingual. Outside the Gaeltacht, English is spoken overall. However, Irish is recognised as the first national language, and is a compulsory subject in both primary and second level schools. There is a growing grassroots demand for Irish-medium education and quite a few families, particularly in urban areas, are raising their children with Irish.

The problem addressed by Mol an Óige manifests differently among native speakers and L2 learners. Native speaker children clearly ‘know’ their phonology, but as explained below, due to the orthographic conventions they do not have a conscious awareness of the consonantal contrasts, and this impacts considerably on literacy acquisition. In the case of the L2 learner, the situation is more complicated: they have neither the sounds in question nor an awareness of there being a contrast of consonantal sounds. The tendency is to simply use the sounds of English – bypassing the language’s phonology, something that affects pronunciation, literacy acquisition, and even acquisition of certain grammatical contrasts (more below). The platform is designed for all cohorts, but the initial step of developing phonological awareness has somewhat different goals for native speaker and L2 learners, and is discussed below.

Efforts are made to make it a fun experience, and to create new materials that subliminally tap into the wider cultural dimensions with songs and stories that have echoes of local traditions and lore. The title Mol an Óige ‘Praise the Young’ comes from the saying mol an Óige agus ticofaidh si (praise the young and they will come with you). As the word mol also means ‘hub’, it captures the aspiration that this platform will be a hub of future resources for early language learning.
Essential prerequisites of literacy acquisition are a grasp of (i) the sound contrasts of the language (an essential part of phonological awareness) and of (ii) how the sounds map to written letters (phonemic awareness). The transparency of the sound-to-letter mapping varies from language to language. When simple and consistent, as in Spanish, Welsh or Finnish, children learn to read much more quickly than when it is complex or irregular, as in English [1].

2.1. The Irish sound system

Irish has a rich sound system, with a contrast of velarised and palatalised consonants (Table 1) not found in English, e.g., initial consonants /l/ and /n/ in the words /l/ο: n/ 'lunch' and /λο:n/ 'lion' [2]. Note that neither of these two lateral consonants are the same as the English /l/, as in /lο:n/ 'lone'. The consonant quality can exert a strong influence on the neighbouring vowels: so for example, when a palatalised consonant abuts a back vowel, as in /λο:n/, a palatal glide may be heard between them. The extent to which such glides are audible varies with dialect.

<table>
<thead>
<tr>
<th></th>
<th>LARYNGAL</th>
<th>DENTAL</th>
<th>ALVEOLAR</th>
<th>ALVEO-LABIAL</th>
<th>PALATAL</th>
<th>VELAR</th>
<th>GLOTTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLOSIVE</td>
<td>p</td>
<td>b</td>
<td>t</td>
<td>d</td>
<td>k</td>
<td>g</td>
<td></td>
</tr>
<tr>
<td>FRICATIVE/ AFFRIQUEANT</td>
<td>f</td>
<td>w</td>
<td>θ</td>
<td>ñ</td>
<td>j</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>NASAL</td>
<td>m</td>
<td>m̃</td>
<td>ñ</td>
<td>n</td>
<td>ñ</td>
<td>ñ</td>
<td></td>
</tr>
<tr>
<td>LATERAL</td>
<td>l</td>
<td>l̃</td>
<td>l̃</td>
<td>l</td>
<td>l̃</td>
<td>l̃</td>
<td></td>
</tr>
</tbody>
</table>

The palatalised-velarised contrast differentiates pairs of words as in the above example. It can also serve to differentiate grammatical meanings: for example, the alternation of the final consonant /n/ and /w/ in the word for 'lion', /λο:n/ → /λο:w/), changes the case from nominative → genitive (see [2]).

2.2. The orthographic conventions

The mapping of sounds to letters in Irish (the phonics system) is also very different from English, being much more opaque though more regular. The complex sound-to-letter mapping of Irish stems partly from the fact that the (Roman) alphabet is a poor fit to the system, in that the contrast involving consonantal quality is not notated. Consonant quality is indirectly signalled through the choice of neighbouring vowel letters. An adjacent 'front' vowel letter “i” or “e” indicates that the consonant is palatalised; a ‘back’ vowel letter indicates it as velarised. Thus, the orthographic representations of the words above are: /λο:n/ = /l/ο: n/ 'lunch' and /leon/ = /l/ο:n/ 'lion' (nominative case); /λοin = /l/ο: n/ 'lunch' and /lein/ = /l/ο:n/ 'lion' (genitive case).

This orthographic convention results in what may look like bewildering sequences of vowel letters in many words. The permutations and combinations of vowel letters that can be used to symbolise consonant quality mean that words with the same vowel phoneme can emerge with different sequences of written vowel letters: Table 2 presents a non-exhaustive list of monosyllabic words with a /i:/ nucleus.

Although the system is opaque, the mapping of sounds to letters in Irish is fairly regular, and for this reason a phonics-based approach to literacy training should be optimal. In spite of this, whole-word memorisation appears to be the mainstream approach [3, 4]. This is far from ideal: Irish is a highly inflected language and words appear in many different forms, making whole-word memorisation an arduous task, compared to what would pertain, say, in English, where there are few inflections.

As learners of Irish are simultaneously learning to read and write English, the large differences in the sound and spelling systems of the two languages greatly complicates matters. Widespread use of the English alphabet, pronounced as in English, and a frequent assumption among teachers that sounds of English can simply be ‘transferred’ from English to Irish, are misguided, and serve rather to obscure the native contrasts of Irish and undermine literacy acquisition.

Despite a growing awareness of the need for a phonologically informed approach [4] there are few resources available to learners and their teachers to develop phonological awareness, the bedrock for developing phonic awareness. With some notable exceptions, such as innovative materials in [5, 6], directed at native-speaker Gaeltacht children in Connacht, current teaching materials are modelled on English and are poorly suited to Irish [7]. The situation for learners with dyslexia is particularly dire, there being neither diagnostic tools nor remediation materials [8, 9]. Some recent research has examined the acquisition of phonological awareness and phonetic principles in native- and second-language learners [10-12].

3. The Platform

The Mol and Óige platform (see homepage in Figure 1) introduces the learner to an imaginary Aran island, populated by characters, animals and objects which appear in the stories and songs to exemplify phonological contrasts.
3.1. Step 1: Training phonological awareness

To make learners consciously aware of the contrast of palatalised and velarised consonants, minimal pairs are used, i.e., words, differing in only one sound – in this case, the words’ initial consonants. Learners are introduced to the pair of words with images and auditory prompts: there are no written materials at this point. The minimal pair subsequently features prominently in illustrated short stories and accompanying musical ditties. For example, the story contrasting /l/ and /l/ as in /lvox/ lion ‘lunch’ and /lvox/ leon ‘lion’ is told with the accompanying illustrations shown in Figure 2. It tells of a sleepy lion who wakes up hungry. He spots little Íde and decides she will do nicely for lunch. (Íde features in another story, as the personification of the long vowel /a/, see illustration in Figure 3.) Happily the Óighean Örga (the ‘Golden Girl’, personifying /a/ in yet another story) comes to the rescue and offers the lion the more succulent lollipop for lunch. All ends well. A short song, also prominently featuring lion and lunch helps memorisation of the contrast.

![Figure 1: Mol an Óige homepage with characters featuring in phonological awareness training.](image)

The learners’ ability to identify and discriminate between the members of the contrast is reinforced and tested through ‘games’ e.g., drag-and-drop of lion or lunch images in response to auditory prompts.

![Figure 2: Images illustrating the ‘lion and lunch’ story.](image)

While Step 1 features word-initial contrasts in a single vocalic environment, it is hypothesised that, once consciously acquired, the contrast will be generalised to other contexts. To promote this, the stories contain examples in varying contexts.

Native speaker children who already have acquired the contrasts, should navigate Step 1 with relative ease, as the aim of making them consciously aware of the contrast should be readily achieved. For the L2 learner, learning the sounds that make up the contrast is a more substantial task, and would need considerably more time and effort.

3.2. Step 2: From sounds to letters

Here, orthographic letters are introduced, but differentiated by colour. This colour-coding has already been subliminally used in the colour schema used in the images illustrating the minimal pairs in Step 1. As illustrated in Figure 3: blue = velarised sounds; yellow = palatalised sounds. The letters are presented along with the image used to train the sound: the yellow/slender/velarised /l/ is illustrated with the lion; the blue/broad/velarised /l/ is illustrated with the lunchbox. The ‘front’ and ‘back’ vowel letters are similarly colour coded. Note that all letters are drawn as wide or narrow to further connect to the widespread use of the terms broad and slender to describe the two classes of vowels.

To go from the individual letters to the encoding of syllables, the learner is presented with tiles in which there are consonant letters (colour coded) and single or combined vowels (also colour coded). A series of quizzes require the learner to find the appropriate tiles to represent auditory prompts. When the correct combinations are chosen, the tiles attach to one another like magnets; when there is a mismatch, they bounce back like magnets repelling each other. The match/mismatch is represented in the lower panel of Figure 3 by the even/jagged abutting edges of the tiles. These matching tasks are not too difficult insofar as the colour coding draws attention to how the phonological contrast is expressed. Once games and puzzles have led to consolidation of the orthographic principles, a version where the colour coding is withdrawn can be attempted.

3.3. Step 3: Early Encoding/Decoding

Building on Step 2, the learner is asked to compose real and nonsense words to attach to image prompts. Imaginary animals/monsters/pets are provided for the composition of nonsense words. A selection of tiles is provided for the task, with colour-coded letters as in Step 2 (Figure 3).

From here, the need is for open ended writing and reading tasks and therefore, the ABAIR synthetic voices are used to provide the auditory output corresponding to learner’s compositions and reading material. As in Step 2, errors in concatenating incompatible consonant and vowel letters sequences are rejected by the system. Visual, colour-coded feedback is provided where there are errors in the choice of palatalised or velarised consonants. Similarly, decoding exercises prompt the learner to read simple syllables, progressing to disyllabic words, nonsense and real, guiding the reader to exploit their knowledge of the sounds and of the phonetic principles in decoding.

3.4. Step 4: Early Reading-Writing

This involves the introduction of very simple texts, with illustrations. It goes from single words to short phrases as would feature in early reader materials – but maximally exploiting the materials used in the earlier steps, and gradually expanding the repertoire. The texts are colour coded as in the earlier steps to prompt the reader’s pronunciation. Readers are encouraged to do syllable-based decomposition of new words, drawing on
their implicit awareness of the phonic principles of Irish, and they can access the (synthetic) spoken version through touching the screen. Simple writing tasks involving words and simple phrases are likewise supported by synthetic spoken versions, and again, ‘spelling’ errors can be made visible by colouring them according to the earlier conventions.

4. Implications for future development

This platform is being developed in stages. A prototype of the first two steps for the basic training of phonological and phonics awareness is almost completed for the Connacht dialect, and needs to be extended to the other two main dialects of Irish for which we have synthetic voices. At that point this version will be released for initial testing in schools.

Work on Steps 3 and 4 is underway. In order to do justice to the potential of these stages, close collaboration with experienced teachers is needed, particularly for the early-reading materials of Step 4. Collaboration with teachers is also needed to test the current prototype (and future iterations) in the classroom, to provide feedback for improving the system. It is hoped that testing of the current prototype will begin as soon as the materials are available in the three dialects for which synthetic voices are now available. Note that ABAIR’s TTS facility is being extended to further sub-dialects, and the platform will need to be updated to ensure the locally most appropriate version is made available.

As the word Mol suggests, the platform is intended as a Hub for a growing suite of learning materials. The four steps described here will hopefully provide a foundation for future targeted materials (in steps) that bring the learner through the complex morphophonemics (e.g., the mutation of initial consonants) and grammatical features (such as the cases and the gender differentiation in nouns and case/gender agreement in adjectives) of Irish in simple, intuitive ways, using songs, stories and quizzes for their consolidation. This will require a close collaboration of teachers, linguists and technologists. In the future, the plan is to further provide teachers with speech and visual resources to build customised content to suit educational targets and the interests of specific learner cohorts.

As it stands, the platform is intended for all learners. It is however of particular value for learners with dyslexia, who currently have no supports for Irish literacy acquisition. Future development envisages platform-based assessment materials, as well as inclusion of additional training materials to support learners with dyslexia and their teachers.

An important consideration in the design of this platform is that it allows learner data to be harvested (with their consent) to generate a database for research into the stages of acquisition. This facility also provides the mechanism for evaluating the effectiveness of each step in the platform. For example, evaluation of Step 1, will be done through the addition of quizzes with each episode, featuring new minimal pair materials, to probe learners’ perception and awareness of the contrast. These quizzes will be modelled on the identification and discrimination perception experiments traditionally used in phonetic research and should allow examination of whether a contrast, once acquired, is generalised beyond the limited context of the initial training materials. Evaluation materials will need to probe the specific goals associated with each step.

Automated tabulation, quantification and graphing of responses to the evaluation materials such as the extra quizzes mentioned here should facilitate not only evaluation of the effectiveness of each training step, but also basic research on the stages of acquisition. This same information can be managed to provide feedback in simple visuals for teachers to monitor progress. Ultimately, such visual feedback may be useful also to learners: seeing concrete evidence of one’s progress can motivate a learner to persist with a task.

Teaching and learning applications such as this are prompting the priorities for ABAIR’s future core technical development, as well as opening up avenues for future linguistic research and application development. A priority for core speech technology is to develop TTS voices for children, in the various dialects. ASR for children’s voices is equally a priority: ABAIR’s current ASR system, ÉIST (see also www abaire.ie), is based primarily on adult recordings and is unlikely to be adequate for use with children at this point. The inclusion of recognition would open up many more avenues to support language learning as well as accessibility in the educational sphere. A CAPT facility, capable of classifying children’s productions of the phonological contrasts being trained here will greatly enhance the platform.

A more distant aspiration is to explore the integration of Mol and Óige with the AAC communication device Geabaire (‘Chatterbox’), currently being developed for Irish. (Geabaire is described in a parallel SIGUL 2023 paper [13]). The AAC system enables non-speaking people to compose sentences by selecting a series of images/words, which are concatenated and spoken out via TTS. A simplified ‘mini version’ of this could be desirable, allowing the merging of accessibility applications with generic educational applications, and it would further help promote creative composition in pre- or non-literate children.

This platform is tailored to specific challenges of Irish language learning, but the linguistic concepts and the use of technology should generalise to other languages. Harnessing speech technology can revolutionise language learning, but the central components of any application are language-specific linguistic and pedagogical knowledge: multidisciplinary research is key.

5. Acknowledgements

This work is supported by An Roinn Turasóireachta, Cultúir, Ealaíon, Gaeltachta, Spóirt agus Meáin with funding from the National Lottery as part of An Stráitéis 20 Bliain don Ghaeilge, 2010 – 2030. It is also partially funded by An Chomhairle um Oideachas Gaeltachtá & Gaelscolaíochta (COGG).
6. References


