

# I'm a useful NLP tool – get me out of here

Monica Ward<sup>1</sup>

**Abstract.** Irish is a compulsory subject in Irish schools. However, there are several pedagogical issues with teaching and learning the language. Computer-Assisted Language Learning (CALL) is under-utilised in schools in Ireland and even more so in the case of Irish, as there are very few CALL resources for the language. This paper looks at the lessons learnt from other Natural Language Processing (NLP)/CALL projects, and tries to apply them to build Intelligent CALL (ICALL) resources for Irish. It shows that a focus on the learner needs and smart use of existing resources can produce useful NLP/CALL resources for language learners. Close collaboration between NLP specialists, CALL researchers, linguists, pedagogical specialists and learners is important in order for a project of this type to be successful. Abair is a useful text-to-speech (TTS) synthesiser for Irish that is relatively unknown outside the TTS/Irish language community. This is a pity as it is a high-quality NLP resource with potential to enhance CALL resources. This paper reports on the integration of Abair into a CALL resource for Irish orthography and pronunciation. While it was developed specifically for Irish, the system is modular in design, and could be adopted for other languages. Furthermore, the lessons learnt are applicable to other languages, not just Irish.

**Keywords:** NLP, ICALL, Irish, Abair.

## 1. Introduction

In Ireland, Irish is a compulsory subject in primary and secondary schools and there are complex socio-cultural issues surrounding the language. Irish orthography is an area of difficulty for learners, which causes problems for spelling and reading Irish words. There are pronunciation rules but they are not fully documented. Furthermore, not many teachers are aware of them and therefore cannot point them

---

1. Dublin City University, Ireland; mward@computing.dcu.ie

**How to cite this article:** Ward, M. (2015). I'm a useful NLP tool – get me out of here. In F. Helm, L. Bradley, M. Guarda, & S. Thoušny (Eds), *Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy* (pp. 553-557). Dublin: Research-publishing.net. <http://dx.doi.org/10.14705/rpnet.2015.000392>

out to their students (Hickey & Stenson, 2011). This makes it difficult for students to read and spell in Irish.

There are many reasons why NLP is not used in CALL, including lack of knowledge and lack of suitability of application to CALL. Abair (2015) is an Irish language synthesiser. It is a high quality NLP tool that allows the user to type in a word, phrase or sentence in Irish and to listen to a spoken version of the text. It provides text to speech synthesis in two dialects and in five different speed settings. Abair is an example of an NLP resource that overcomes some of the NLP/CALL integration problems. It is a high quality resource that is based on a theoretically sound design plan with a long-term perspective. It has continued to develop over a period of years and has a long-term vision. An NLP resource such as Abair would be difficult and expensive to build from scratch – but it exists and can be used in ICALL resources. To date, Abair is not well-known outside the inner core of Irish language NLP researchers and has a focus on text to speech synthesis rather than language learning.

There is a need for CALL resources for Irish. Several CALL resources exist, but they are sometimes developed by language enthusiasts (with limited pedagogical experience) or may be technically correct (but not appealing to non-linguistic language learners). The majority of Irish language learners are primary and secondary school children who learn it as part of the core curriculum. In general, they are not taught the pronunciation and orthography rules of the language. Often, their teachers are not aware of the rules themselves. Often, the parents are similarly unaware of the rules. Currently, there are no quality CALL resources for Irish pronunciation and orthography – but there is a definite need for them.

## **2. Method**

### **2.1. Approach**

Rather than starting from scratch with the development of a CALL resource for Irish pronunciation, it made sense to look at what research and resources were already available and to try to utilise them where possible. In this regard, there were three main contributions to consider. Firstly, Hickey and Stenson (2011) provide a suggestion as to how the rules of Irish orthography and pronunciation could be taught to learners. They stress that the analysis of the rules was still under development, but the most basic rules were fairly well-established and could be explained to learners. Hickey is an educational psychologist and is an expert on the teaching and learning of Irish as both a first and a second language. Stenson is

an expert Irish language linguist and they have worked together in recent years on Irish orthography. Secondly, there is the Abair project. Abair is a high-quality TTS synthesiser for Irish, which currently provides TTS for the Donegal and Connemara dialects of Irish. Thirdly, there was the body of research on why NLP was not more widely used in CALL. Researchers over the years have noted that there have been several NLP resources developed that could potentially be useful in CALL resources, but never made the transition from the NLP world to the CALL world. It is easy to see why this might be the case. For example, [Kraif et al. \(2004\)](#) identified three reasons for this: NLP techniques may not be reliable, NLP resources are difficult and expensive to implement, and end-users may not be aware of NLP possibilities. [Granger et al. \(2006\)](#) note that the lack of communication between NLP specialists, CALL researchers and language teachers is a major challenge to NLP/CALL integration. Many ICALL projects disappear due to lack of funding, lack of long-term perspective, and lack of understanding of pedagogical issues ([Tschichold, 2014](#)).

Learning from this research, it seemed logical to use [Hickey and Stenson's \(2011\)](#) work, combined with Abair, to produce a CALL resource for Irish orthography and pronunciation. This combination overcomes some of the NLP/CALL integration difficulties outlined above. Developing a CALL resource that utilised Abair would make its technology available to end-users (without them having to be aware of the technicalities). Pedagogical and linguistic input from Irish language specialists would help to overcome some of the ICALL/pedagogical issues that often arise.

## 2.2. CALLIPSO system

The CALLIPSO (CALL for Irish for Parents, Students and Others) system was designed and developed as a CALL resource for Irish orthography and pronunciation. It was based on [Hickey and Stenson's \(2011\)](#) work and uses TTS outputs from the Abair system. The initial version was designed with parents in mind, but it is a modular system and is designed to be customisable to the learning needs of the end-user. This is an overlooked group in terms of CALL resources for Irish. Many parents want to be able to help their children with homework, particularly in primary school. However, they have often forgotten their Irish or may not have a particularly good understanding of the language. Furthermore, in recent years, Ireland has seen an increase in immigrant parents who have no previous Irish language experience and there are no accessible resources for them. In most cases, the parents just want to be able to check their children's spelling and reading, without actually having to learn the language. Their language needs are different from the traditional language learner.

The CALLIPSO system provides information on Irish consonants and vowels, along with sample sounds and words (provided by Abair). It provides some pronunciation exercises for learners so they can check their understanding. It is currently aimed at (false) beginner learners and more detailed information will be provided for more advanced learners (e.g. trainee or current teachers) in future versions. The CALLIPSO system is designed to be L1 and target language independent, although obviously, the target language is Irish in this case. It is planned to provide a version with Polish as the L1, as this is one of the major immigrant languages in Ireland at this time.

### **3. Discussion and conclusions**

It would not have been possible to design and develop the CALLIPSO without reusing existing NLP and research resources. Abair is an example of a high-quality NLP resource that could and should be used more outside its current environs. The initial version of the CALLIPSO system did not have access to the Abair application program interface (API), but it will use the API in future versions. By focusing on learner needs (i.e. the need to be able to understand Irish orthography and pronunciation), it was possible to develop a CALL resource that would be useful to the target learner group. Feedback from parents has been positive and there are plans for further improvements to the system.

As a final remark, combining research with an NLP tool (Abair) made it possible to build a CALL resource that was based on pedagogical guidelines and that was of benefit to the target learner group. The useful NLP tool was able to escape into the wild.

### **4. Acknowledgements**

We would like to thank the abair.ie team in Trinity College Dublin for access to the Abair system. Also thanks to the Centre for Digital Content Platform Research (ADAPT), the global centre of excellence for digital content and media innovation, for its support.

### **References**

- Abair. (2015). *Abair.ie – The Irish Language Synthesiser*. Retrieved from <http://www.abair.tcd.ie/>
- Granger, S., Antoniadis, G., Fairon, C., Medori, J., & Zampa, V. (2006). *Report on NLP-based CALL workshop*. Research report - Report number D39.3.1. 2006. Retrieved from <https://telearn.archives-ouvertes.fr/hal-00190372/document>

- Hickey, T., & Stenson, N. (2011). Irish orthography: what do teachers and learners need to know about it, and why? *Language, Culture and Curriculum*, 24(1), 23-46. doi:10.1080/07908318.2010.527347
- Kraif, O., Antoniadis, G., Echinard, S., Loiseau, M., Lebarbé, T., & Ponton, C. (2004). *NLP tools for CALL: the simpler, the better*. In InSTIL/ICALL Symposium 2004. Retrieved from [http://www.cs.columbia.edu/~amaxwell/candidacy/l2learning/iic4\\_009.pdf](http://www.cs.columbia.edu/~amaxwell/candidacy/l2learning/iic4_009.pdf)
- Tschichold, C. (2014). *Challenges for ICALL*. Keynote speech, 2nd workshop on NLP for computer-assisted language learning NoDaLiDa workshop, May 22, 2013, Oslo, Norway. Retrieved from [http://spraakbanken.gu.se/sites/spraakbanken.gu.se/files/ICALL\\_handout\\_invited\\_talk.pdf](http://spraakbanken.gu.se/sites/spraakbanken.gu.se/files/ICALL_handout_invited_talk.pdf)

Published by Research-publishing.net, not-for-profit association  
Dublin, Ireland; info@research-publishing.net

© 2015 by Research-publishing.net (collective work)  
© 2015 by Author (individual work)

Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy  
Edited by Francesca Helm, Linda Bradley, Marta Guarda, and Sylvie Thouéšny

**Rights:** All articles in this collection are published under the Attribution-NonCommercial -NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Under this licence, the contents are freely available online (as PDF files) for anybody to read, download, copy, and redistribute provided that the author(s), editorial team, and publisher are properly cited. Commercial use and derivative works are, however, not permitted.



**Disclaimer:** Research-publishing.net does not take any responsibility for the content of the pages written by the authors of this book. The authors have recognised that the work described was not published before, or that it is not under consideration for publication elsewhere. While the information in this book are believed to be true and accurate on the date of its going to press, neither the editorial team, nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, expressed or implied, with respect to the material contained herein. While Research-publishing.net is committed to publishing works of integrity, the words are the authors' alone.

**Trademark notice:** product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

**Copyrighted material:** every effort has been made by the editorial team to trace copyright holders and to obtain their permission for the use of copyrighted material in this book. In the event of errors or omissions, please notify the publisher of any corrections that will need to be incorporated in future editions of this book.

Typeset by Research-publishing.net  
Fonts used are licensed under a SIL Open Font License

ISBN13: 978-1-908416-28-5 (Paperback - Print on demand, black and white)  
Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

ISBN13: 978-1-908416-29-2 (Ebook, PDF, colour)  
ISBN13: 978-1-908416-30-8 (Ebook, EPUB, colour)

Legal deposit, Ireland: The National Library of Ireland, The Library of Trinity College, The Library of the University of Limerick, The Library of Dublin City University, The Library of NUI Cork, The Library of NUI Maynooth, The Library of University College Dublin, The Library of NUI Galway.

Legal deposit, United Kingdom: The British Library.  
British Library Cataloguing-in-Publication Data.  
A cataloguing record for this book is available from the British Library.

Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: décembre 2015.