Check for updates

Joining the blocks together – an NLP pipeline for CALL development

Monica Ward[1]

Abstract. Intelligent Computer-Assisted Language Learning (ICALL) involves using tools and techniques from computational linguistics and Natural Language Processing (NLP) in the language learning process. It is an inherently complex endeavour and is multi-, inter-, and trans-disciplinary in nature. Often these tools and techniques are designed for tasks and purposes other than language learning, and this makes their adaptation and use in the CALL domain difficult. It can be even more challenging for Less-Resourced Languages (LRLs) for CALL researchers to adapt or incorporate NLP into CALL artefacts. This paper reports on how two existing NLP resources for Irish, a morphological analyser and a parser, were used to develop an app for Irish. The app, Irish Word Bricks (IWB), was adapted from an existing CALL app – Word Bricks (Mozgovoy & Efimov, 2013). Without this 'joining the blocks together' approach, the development of the IWB app would certainly have taken longer, may not have been as efficient or effective, and may not even have been accomplished at all.

Keywords: NLP, Irish, ICALL, Word Bricks, young learners.

1. Introduction

The development of CALL resources is challenging (Godwin-Jones, 2015) and it can be difficult to incorporate NLP technologies in CALL resources (Heift & Schulze, 2007). Technical challenges arise from the fact that the NLP tools may focus on language from a specified domain or linguistic standard. Cross-domain challenges relate to different foci and research aims between NLP and CALL researchers. Research that is focussed on building a robust NLP tool might be different to one that is suitable for foreign language learners. Another problem

1. Dublin City University, Dublin, Ireland; monica.ward@dcu.ie; https://orcid.org/0000-0001-7327-1395

How to cite this article: Ward, M. (2019). Joining the blocks together – an NLP pipeline for CALL development. In F. Meunier, J. Van de Vyver, L. Bradley & S. Thouësny (Eds), *CALL and complexity – short papers from EUROCALL 2019* (pp. 397-401). Research-publishing.net. https://doi.org/10.14705/rpnet.2019.38.1043

© 2019 Monica Ward (CC BY)

that can arise is that researchers from both domains may not really understand one another. NLP researchers may shy away from working with language learning and language teachers may be afraid of the complexities of NLP.

Natural languages are also complex, nuanced, and ambiguous, and this makes NLP challenging. NLP is a broad field and includes text to speech technology, grammar checkers, machine translation, and artificial intelligence chatbots, to name but a few. Some NLP tools are CALL ready. Text to speech technology is ready for use in L2 classrooms (Cardoso, Smith, & Garcia Fuentes, 2015), and automatic corrective feedback from grammar checkers have been shown to be beneficial (Ferris, Liu, Sinha, & Senna, 2013).

Often, ICALL artefacts will be built using already existing NLP tools. However, not all languages have the same level of resources. This makes developing new NLP tools for LRLs more challenging, and researchers have to be creative in how they work. This paper reports on how two NLP tools (a morphological analyser and a parser) were joined together to produce an NLP pipeline for an app designed for Irish, a LRL and a Less Commonly Taught Language (LCTL).

2. Context

2.1. Irish

Irish is a compulsory subject in schools in the Republic of Ireland. There are very few CALL resources for Irish and most of these are research-based and not widely used or are rather rudimentary in nature. It is not a major language, there is no commercial incentive to develop NLP resources for the language, and the pool of experts and researchers in the field is limited. These factors make it difficult to develop NLP resources for Irish.

2.2. Word Bricks

Word Bricks was originally developed for Japanese university students learning English (Mozgovoy & Efimov, 2013). It is based on a visual learning paradigm, facilitates an active learning (Prince, 2004), 'hands on' approach (Dewey, 1997), and channels the idea that perception precedes production (Flege, 1995). With Word Bricks, learners can experiment with the sentence structure of English. They can combine bricks (parts of speech) of different colours and shapes together to

construct sentences and only grammatically correct sentences are possible. One of the motivations behind Word Bricks is to overcome the problem of students being limited to generic homework activities (Howard & Major, 2004) by enabling learners to play around with parts of speech. Word Bricks taps into the mobile learning 'game' idea which can increase learner motivation (Ducate & Lomicka, 2009).

2.3. IWB

The IWB app was derived from the original Word Bricks. The target learner group was primary school children in a classroom context. The initial phase (Phase 0) was an exploratory phase in which the feasibility of developing the IWB app was investigated. This phase was quite manual but demonstrated that the approach was feasible.

In Phase 1, the initial version of the IWB app was designed and developed with a team that included teachers, Irish CALL researchers and the original IWB team. A user-centred approach was used whereby the teachers were consulted before the development of content and throughout the process. Two existing Irish NLP resources – a morphological analyser (Uí Dhonnchadha, 2002) and a treebank and parser (Lynn, 2016) – were used manually as stand-alone resources to confirm the correctness of the information sent to the Word Bricks team. By adapting the existing Word Bricks app (with the original Word Bricks team) that had proved to be successful (Park, Purgina, & Mozgovoy, 2016), the authors were able to develop an interactive app for Irish that would otherwise have taken much longer to develop (if at all).

One limitation of the first version of the app was that learners could only use words from a pre-defined vocabulary list. In Phase 2 of the IWB app, the process was automated. The IWB team were able to integrate the NLP tools into the IWB app engine. This meant that the app can now process sentences with user-inputted words and gives the learners much more freedom and an improved platform for experimentation. Future research will focus on evaluating the new version of the IWB app.

3. Discussion

The IWB app has been used successfully by learners in two Irish primary schools (Ward, Mozgovoy, & Purgina, 2019). Students from the ages of seven to 12 have

successfully used the app to revise five different sentence structures (possession, feelings, actions, locations, and questions). Students and teachers were surveyed after using the IWB app and the feedback was positive (Purgina, Mozgovoy, & Ward, 2017). A similar continuum to the mobile assisted language learning continuum (Barcomb, Grimshaw, & Cardoso, 2018) could be used in the ICALL context, with the observation that it is probably more challenging for teachers to move from adaptation through to creation in the ICALL domain. Without the collaboration between researchers with NLP expertise, CALL expertise, and teaching expertise combination, the non-NLP researcher would not have been able to develop the app alone, while the NLP experts would not have been able to envision how their NLP tools could be adapted for the primary school context. The development of the enhanced IWB app (Phase 2) indicates the success of the pipeline approach. The original NLP technologies were developed with machine translation in mind. However, with creative re-imaging of their purpose, they were able to be joined together to provide an NLP pipeline for the IWB app.

4. Conclusion

Just as technology should not be used for technology's sake, NLP should not be used 'just because it exists'. Pedagogical issues are primordial and this is why the user-centred approach was a key feature of the app development. The CALL community should learn from others and not reinvent the wheel (Gimeno-Sanz, Sevilla-Pavón, & Martínez-Sáez, 2018). This is especially important in the LCTL community where lessons can be learnt from 'bigger' languages. The IWB team learnt from the original Word Bricks team and were able to leverage their experience and expertise to develop the IWB app.

References

Barcomb, M., Grimshaw, J., & Cardoso, W. (2018). Instructors as MALL engineers: adapting, modifying, and creating mobile materials for listening practice. In P. Taalas, J. Jalkanen, L. Bradley & S. Thouësny (Eds), *Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018* (pp. 11-15). Research-publishing.net. https://doi.org/10.14705/rpnet.2018.26.805

Cardoso, W., Smith, G., & Garcia Fuentes, C. (2015). Evaluating text-to-speech synthesizers. In F. Helm, L. Bradley, M. Guarda & S. Thouësny (Eds), *Critical CALL Proceedings of the 2015 EUROCALL Conference, Padova, Italy* (pp.108-113). Research-publishing.net. https://doi.org/10.14705/rpnet.2015.000318

Dewey, J. (1997). *Democracy and education: an introduction to the philosophy of education.* Free Press [First published 1916].

Ducate, L., & Lomicka, L. (2009). Podcasting: an effective tool for honing language students' pronunciation? *Language Learning & Technology, 13*(3), 66-86.

Ferris, D., Liu, H., Sinha, A., & Senna, M. (2013). Written corrective feedback for individual L2 writers. *Journal of Second Language Writing, 22*(3), 307-329. https://doi.org/10.1016/j.jslw.2012.09.009

Flege, J. (1995). Second language speech learning: theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 233-272). York Press.

Gimeno-Sanz, A., Sevilla-Pavón, A., & Martínez-Sáez, A. (2018). From local to massive learning: unveiling the (re) design process of an English LMOOC based on InGenio materials. In P. Taalas, J. Jalkanen, L. Bradley & S. Thouësny (Eds), *Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018* (pp. 77-82). Research-publishing.net. https://doi.org/10.14705/rpnet.2018.26.816

Godwin-Jones, R. (2015). Emerging technologies the evolving roles of language teachers: trained coders, local researchers, global citizens. *Learning and Technology, 19*(1), 10-22.

Heift, T., & Schulze, M. (2007). *Errors and intelligence in computer-assisted language learning: Parsers and pedagogues*. Routledge. https://doi.org/10.1017/s0958344009000044

Howard, J., & Major, J. (2004). Guidelines for designing effective English language teaching materials. *The TESOLANZ Journal, 12*(10), 50-58.

Lynn, T. (2016). *Irish dependency treebanking and parsing*. Doctoral dissertation, Dublin City University.

Mozgovoy, M., & Efimov, R. (2013). WordBricks: a virtual language lab inspired by Scratch environment and dependency grammars. *Humancentric Computing and Information Sciences, 3*(1), 1-9. https://doi.org/10.1186/2192-1962-3-5

Park, M., Purgina, M., & Mozgovoy, M. (2016). Learning English Grammar with WordBricks: classroom experience. In *Proceedings of the 2016 IEEE International Conference on Teaching and Learning in Education*.

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education, 93*(3), 223-231. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x

Purgina, M., Mozgovoy, M., & Ward, M. (2017). MALL with WordBricks–building correct sentences brick by brick. In K. Borthwick, L. Bradley & S. Thouësny (Eds), *CALL in a climate of change: adapting to turbulent global conditions – short papers from EUROCALL 2017* (pp. 254-259). Research-publishing.net. https://doi.org/10.14705/rpnet.2017.eurocall2017.722

Uí Dhonnchadha, E. (2002). *An analyser and generator for Irish inflectional morphology using finite-state transducers*. Doctoral dissertation, Dublin City University.

Ward, M., Mozgovoy, M., & Purgina, M. (2019, forthcoming). A green Approach for an Irish app (refactor, reuse and keeping it real). *Celtic Language Technology Workshop, MT Summit XVII,* Dublin City University, August 2019.

Published by Research-publishing.net, a not-for-profit association
Contact: info@research-publishing.net

© 2019 by Editors (collective work)
© 2019 by Authors (individual work)

CALL and complexity – short papers from EUROCALL 2019
Edited by Fanny Meunier, Julie Van de Vyver, Linda Bradley, and Sylvie Thouësny

Publication date: 2019/12/09

Rights: the whole volume is published under the Attribution-NonCommercial-NoDerivatives International (CC BY-NC-ND) licence; **individual articles may have a different licence**. Under the CC BY-NC-ND licence, the volume is freely available online (https://doi.org/10.14705/rpnet.2019.38.9782490057542) 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 was not under consideration for publication elsewhere. While the information in this book is 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. 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
Cover theme by © 2019 Frédéric Verolleman
Cover layout by © 2019 Raphaël Savina (raphael@savina.net)

Fonts used are licensed under a SIL Open Font License

ISBN13: 978-2-490057-54-2 (Ebook, PDF, colour)
ISBN13: 978-2-490057-55-9 (Ebook, EPUB, colour)
ISBN13: 978-2-490057-53-5 (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'.

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 2019.