Joining the blocks together – an NLP pipeline for CALL development

Monica Ward

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
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
Monica Ward

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


