Using Irish NLP resources in Primary School Education

Monica Ward
School of Computing
Dublin City University
Ireland
mward@computing.dcu.ie

Abstract

This paper looks at the use of Natural Language Processing (NLP) resources in primary school education in Ireland. It shows how two Irish NLP resources, the Irish Finite State Transducer Morphological Engine (IFSTME) (Uí Dhonnchadha, 2002) and *Gramadóir* (Scannell, 2005) were used as the underlying engines for two Computer Assisted Language Learning (CALL) resources for Irish. The IFSTME was used to supply verb conjugation information for a Verb Checker Component of a CALL resource, while *Gramadóir* was the underlying engine for a Writing Checker Component. The paper outlines the motivation behind the development of these resources which include trying to leverage some of the benefits of CALL for students studying Irish in primary school. In order to develop CALL materials that were not just an electronic form of a textbook, it was considered important to incorporate existing NLP resources into the CALL materials. This would have the benefit of not re-inventing the wheel and of using tools that had been designed and testing by a knowledgeable NLP researcher, rather than starting from scratch. The paper reports on the successful development of the CALL resources and some positive feedback from students and teachers. There are several non-technical reasons, mainly logistical, which hinder the deployment of Irish CALL resources in schools, but Irish NLP researchers should strive to disseminate their research and findings to a wider audience than usual, if they wish others to benefit from their work.

1 Introduction

This paper looks at how Irish NLP resources can be used in the development of Computer Assisted Language Learning (CALL) resources. It reports on the motivation for using CALL and specifically NLP/CALL in the primary school context in Ireland. Irish is a compulsory subject in primary schools in Ireland and most students spend 13 years studying the language (Murtagh, 2003), but it is not a particularly popular subject (Ó Riagáin and Ó Glíasáin, 1994, DCRGA, 2009) . CALL has many potential benefits for the language learner and it is important the students learning Irish have access to reliable, good quality CALL resources. However, it is difficult to develop such CALL resources, as usually a multi-disciplinary team is required, and such a team is often hard to assemble. One approach is to try to adapt and reuse existing resources to speed up the development process and indeed, provide resources that might not otherwise exist.

With this in mind, two existing NLP resources for Irish were used to develop CALL resources for students in the primary school context. The use of the resources is not limited to primary school students, but they were developed with these students as the target learning group. The first tool that was used was the Irish Finite State Transducer Morphology Engine (Uí Dhonnchadha, 2002). It was used to provide verb conjugation information for the Verb Conjugation Component (VCC) of the CALL resources. The aim of the VCC was to provide static and dynamic web pages with verb conjugation information and exercises/language games for the learner. The second tool used was *Gramadóir* (Scannell, 2005). It is a grammar checking tool and provided the underlying engine for the Writing Checker Component for the CALL resources. A wrapper was placed around *Gramadóir* in order to adapt it for the target learners. This included modifying the errors messages to be more young-learner friendly and separating spelling and grammar errors. CALL resources were developed using these

Proceedings of the First Celtic Language Technology Workshop, pages 6–17,
Dublin, Ireland, August 23 2014.

Irish NLP resources and deployed in two primary schools in Ireland. The students were able to use the resources without any major difficulties, but long term use depends on factors other than the NLP/CALL integration ones. However, in order to make use of the NLP resources that are currently available to CALL developers, it behoves NLP researchers to make their research widely available and comprehensible to a non-NLP knowledgeable audience. Of course, CALL researchers should also try to interact with the NLP community for a fruitful exchange of ideas and knowledge.

2 Background

Irish used to be the lingua franca in Ireland many centuries ago, but this is no longer the case. However, the vast majority of school students in Ireland study Irish for 13 years (Murtagh, 2003) in both primary and secondary school. There are several challenges to the teaching of Irish, including attitude, potential pedagogical difficulties and lack of suitable resources (including computer-based resources). This section looks at the place of Irish in the primary school system in Ireland, the problem of lack of suitable, high-quality, reliable resources for Irish for learners in general and especially for primary school children. It also looks at the role of Natural Language Processing (NLP) and Computer Assisted Language Learning (CALL) in the teaching and learning of Irish.

2.1 Irish

Irish is a morphologically-rich language that was the lingua-franca of the majority of people in Ireland until around the 17th century. Its use started to decline around this time and today there are approximately 20,000 active speakers (Ó hÉallaithe, 2004). Irish has had a complex, paradoxical socio-cultural role in Ireland. On the one hand, people in Ireland appreciate the importance of having a national language that is distinct to Ireland and understand its cultural role (DCRGA, 2009). However, they are somewhat ambivalent about its role in the education system.

2.2 Education

There are several pedagogical issues with the teaching of Irish in schools in Ireland. It is one of the core subjects and is taught on a daily basis. Often there is a lack of interest on the part of the students and their parents. Reasons such as 'it's a useless language, no one speaks it anymore', or 'why don't they teach French/Chinese instead?' are sometimes heard. Some students find it difficult. Eleven of the most commonly used verbs are highly irregular, which can be daunting and confusing for young learners. There is also the issue with lack of resources. Obviously, there is no large international market for Irish language primary school text books and publishers only have the internal market in Ireland. This limits the financial incentive for publishers to provide materials for students. In many primary schools, students have to pay for their own books, with some schools operating book rental schemes. This means that for any schools there is little or no incentive to change the books series that they use for teaching Irish. Furthermore, given the non-positive attitude some parents have towards the time/effort devoted to teaching and learning Irish in primary school, they are often not receptive to moving to a different book series if they do not have the option to buy pre-owned books for older children in the school. Harris and Murtagh (1999) and Hickey and Stenson (2010) provide a good overview of the Irish education field.

2.3 Lack of Suitable Resources

One possible strategy to incorporate a more modern approach is to use electronic resources. However, many of the resources available are not particularly suitable for primary schools students, as they are aimed at adults or may not be very accurate. Adults may be able to comprehend that the information that they see online may not be totally correct, but primary school students are not accustomed to this, as they expect the information to be correct all the time. For example, an adult may understand that "The President has super powers" or "London is the capital of Ireland" may not be true, but a child may just accept it as fact.

2.4 NLP, Computer Assisted Language Learning and Irish

Computer Assisted Language Learning (CALL) can help in the language learning process. It can help with learner motivation (e.g. Murphy and Hurd, 2011) and provide a degree of privacy for students. It enables students to repeat exercises and revise as often as they like – the computer will not tire of providing feedback to students (unlike, perhaps, a teacher in a classroom setting). Students can work at their own pace when using CALL resources – something which can be helpful in a mixed-ability class. CALL can be useful when there is limited or no access to a teacher e.g. in a minority or endangered language scenario. CALL can perhaps enhance the prestige of a minority language, by demonstrating that the language as an electronic and/or online presence. All these potential benefits can accrue to CALL for Irish. The problem is that there are several issues which hinder the development and deployment of CALL resources for Irish. From a CALL resource development point of view, the teachers may not have the time, knowledge or the expertise to develop CALL materials. There may not be the computing resources for the students to have access to the CALL materials. These factors pertain for Irish in the primary school context. The teachers cover all primary school subjects and, in general, are not trained linguists or Irish language specialists. Furthermore, while they may have reasonable computing skill, they may not have the skills and knowledge necessary required to develop Irish CALL materials. In many primary schools in Ireland, there may not be a computer in the classroom and so the students have to use a computer lab. Often, the computers are relatively old and are of a low specification, and the students have limited access to the lab. In their weekly computer slot, the teacher has to decide to use the time for English, mathematics or other school subjects.

Many CALL resources do not use any NLP e.g. the BBC Languages (World Service English) (BBC, 2014) is a general CALL resource for English language learners. Intelligent CALL (ICALL) mainly draws on Natural Language Processing (NLP) and Intelligent Tutoring Systems (ITS) (Matthews, 1993). NLP technologies can be used in CALL resources for concordancing, morphological processing and syntactic processing (Nerbonne, 2003). There are many reasons why NLP technologies are not widely used in CALL. NLP is inherently difficult and there are difficulties in integrating NLP in CALL resources. NLP researchers and NLP research is not CALL-based and there are difficulties in visualising how NLP can be used in CALL resources. Furthermore, there is a lack of knowledge amongst CALL practitioners about NLP, as the use of NLP in CALL has been driven by NLP specialists rather than CALL practitioners. Another difficulty is that NLP tools and techniques are often designed to work with correct input (Vandeventer Faltin, 2003) and language learners produce incorrect input. Also, some NLP CALL projects concentrate on the functionality/content and neglect the User Interface (UI) and this makes it difficult for the non-expert user to use the resources. However, there is a growing interest in NLP resources for language learners, particularly in the area of error detection (Leacock et al., 2014). There have been some successful NLP CALL programs (e.g. ALICE-chan (Levin and Evans, 1995)), but there are not many good examples that demonstrate the ability of NLP in CALL. Many NLP/CALL projects finish at the prototype stage and progress no further. The issue of using NLP in CALL without a good pedagogical basis must also be noted. There are also some socio-cultural factors that must also be considered including the attitudes of teachers, learners and NLP researchers to the NLP/CALL field. There are very few NLP resources available for Irish. However, two of these resources, the IFSTE and *Gramadóir*, are robust and informative and can be used in CALL resources for Irish and these are discussed below.

3 Resources

3.1 Approach

As outlined above, there is a problem with the lack of suitable, high quality CALL resources for Irish. One potential solution to this problem is to use existing NLP resources for Irish in CALL resources for the language. There are not too many such resources available for Irish, but two very useful resources are *Gramadóir* (Scannell, 2005) and the Irish Finite State Transducer Morphology Engine (Uí Dhonnchadha, 2002) (henceforth, IFSTME). These are both high-quality, reliable and accurate resources that are publicly available. These resources were integrated into two Irish CALL resources for primary school children. *Gramadóir* was used in a Writing Checker Component (WCC) and the IFSTME was used in a Verb Conjugation Component (VCC). The overall architecture ran on an Apache

server, with static pages stored in the `htdocs` directory and dynamic pages stored in the `cgi-bin` directory. XML technologies and Perl were core components of the CALL software.

3.2 Verb Conjugation Component

Uí Dhonnchadh's (2002) Irish Finite State Transducer Morphology Engine (IFSTME) is a comprehensive resource which supplies morphological information for Irish words and sentences. The IFSTME was used to generate the verb conjugations for verbs in the past simple tense.

The aim of the Verb Conjugation Component (VCC) is to provide a tool to produce static and animated verb conjugation web pages based on externally supplied verb data. The underlying engine is an Irish Finite State Transducer Morphology Engine (IFSTME) (Uí Dhonnchadha, 2002). It was combined with an animation tool (Koller, 2004) and a CALL Template (Ward, 2001) to provide an Irish verb learning tool for primary school students. Figure 1 shows the information flow for the VCC. The external source of verb information (i.e. the IFSTME) provides information on verbs to the VCC which uses the information in the CALL resources.

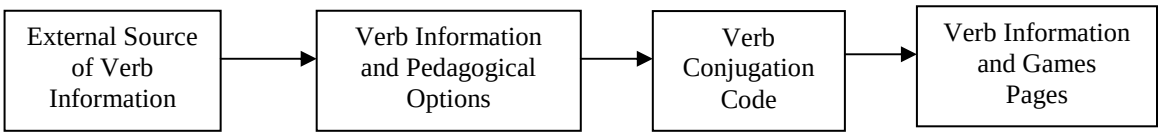

Figure 1 Information Flow for the Verb Conjugation Component

Figure 2 shows an overview of the VCC. The external verb information (from the IFSTME) is combined with local code files and local configuration files in the VCC. The teacher provided pedagogical input to the process. The VCC combines this data with flash animation code to produce verb information files, activity files and report files for the learner to use. The teacher can also see the report files.

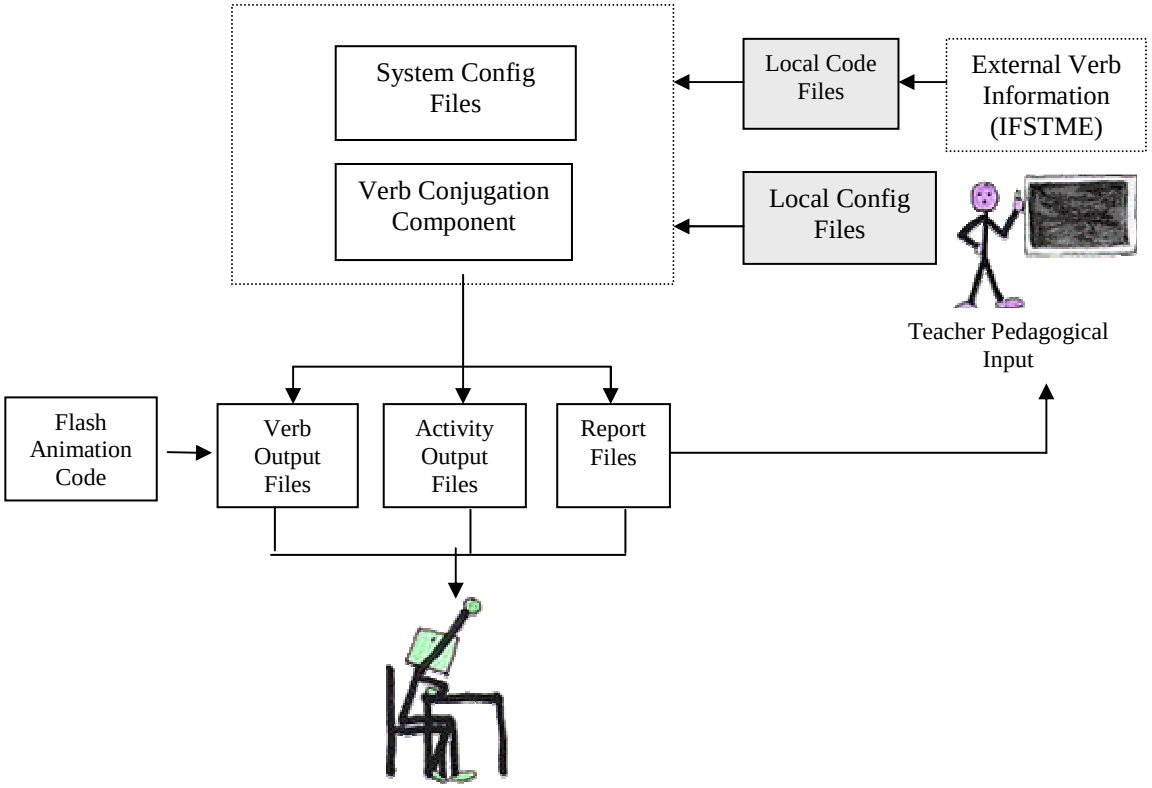

Figure 2: Overview of the VCC

The IFSTME provides an analyser and generator for Irish inflectional morphology for nouns, adjectives and verbs. Replace rule triggers (for stems and affixes) are combined with replace rules written as regular expressions (for word mutations) to produce a two-level morphological transducer for Irish. The VCC only uses a very small subset of the verb forms provided by the IME (there are 52 forms in all). It has web pages for 20 verbs, in both static and dynamic forms. Figure 3 shows the past indicative information for *bris* (to break) supplied by the Irish Finite State Morphology Engine (Uí Dhonnchadha, 2002). Note that the output is not intended to be used as presented by the end-user, hence the presence of ^FH and ^FS tags in Figure 3. Figure 4 shows the animated verb page for *bris* (past tense).

Bris+Verb+PastInd	b^FHris
Bris+Verb+PastInd+1P+Pl	b^FHris^FSeamar
Bris+Verb+PastInd+Auto	bris^FSeadh
Bris+Verb+PastInd+Auto+Neg	bris^FSeadh
Bris+Verb+PastInd+Auto+NegQ	bris^FSeadh
Bris+Verb+PastInd+Auto+Q	bris^FSeadh

Figure 3: Past Indicative Information for *bris* (to break)

Figure 4: Animated Verb Page for *bris*

3.3 Writing Checker Component

The Writing Checker Component (WCC) provides a tool that checks the learner's text input and provides feedback on spelling and grammar errors. It adapts an externally supplied grammar checker, *Gramadóir* (Scannell, 2005) to the needs of primary school students. *Gramadóir* is an open source grammar checker that has been implemented for Irish and it can be used on a variety of operating systems. It is modular in design and provides separate components for sentence segmentation, spell checking, part-of-speech tagging and grammar checking. It is easy to use and there is a simple command line interface and a web interface to the software. It is corpus-based and is booted from web-based corpora. It is easy to port to other languages as the language developers' pack provided is designed so that no programming experience is required. It is scalable. Spell checking packages can be developed in a few hours, while the engine also accommodates the development of a full-scale grammar checker.

Gramadóir is an excellent, accurate Irish language resource. It is aimed at linguistically-aware adults. It can be used in white-box mode and be adapted to the needs of the users. However, a black-box approach was taken when developing a writing checker for primary school students. Under this approach, the grammar error messages to the user were passed through a filter and substituted with more suitable error messages for the target learners.

There was an initial pilot study to test the feasibility of the resources and there were several design modifications based on learner and teacher feedback. For example, there was a need to convert the adult learner-oriented language of *Gramadóir*'s errors messages to language more appropriate to younger learners. Some of the original *Gramadóir*'s error messages and their WCC equivalent are shown in Table 1. Note that not all students would understand the words "*urú*" and "*séimhiú*" even thought the teacher may have explained them.

There was a need to separate out spelling errors from grammar errors and an error classification file was used to classify *Gramadóir*'s errors as either grammar or spelling errors. Sometimes, *Gramadóir* failed to suggest any alternatives for spelling errors and the Levenshtein algorithm (implemented with code from Merriampark (2005)) was used to check suitable words from the local dictionary. The local dictionary consisted of words from the some class texts. A word with a Levenshtien value of 1 was probably the word the student intended to use, while those with a value of 2 were probably suitable. There was also a need to be able to correct and resubmit a text. The screen layout had to be changed so that more information could be viewed at once and to minimise scrolling. A review of the errors detected and not detected by *Gramdóir* was required and certain adaptations were necessary.

Gramadóir Message	Writing Checker Message
	Humm, there might be an error here
Definite article required	'an' required
Eclipsis missing	You need a letter at the start of the word
Lenition missing	You might be missing a 'h' here
Prefix \/d'\/ missing	You need a 'd' here
The dependent form of the verb …	The verb is not correct
The genitive case	You need to add something here

Table 1. *Gramadóir* Error Messages and their WCC Equivalent

Table 2 shows some sample student text, along with some of the error types and the changes made to *Gramadóir*'s error messages. Note that the missing word "*seomra*" before "*suite*" was not detected in example 3 in Table 2.

Error Type	Text	*Gramadóir* Error Message	Expected Error	New Error Message
Gramadóir error OK	Tá bosca beag agam ach tá níos bosca lú agat..	Usually used in the set phrase /níos lú, is lú/	As expected	Usually used in the set phrase /níos lú, is lú/
Gramadóir error OK, but msg not suitable	Tá trí gloine atá an mbord.	Unnecessary eclip-sis	As expected	Maybe you should have **ar an mbord**
Error detected, but should be ig-nored	Shuigh Ciara agus Maire sa suite ar an tolg.	It seems unlikely that you intended to use the subjunctive here (Maire)		
Error incorrectly detected	Fuair **Ríona** páipéar.	Unnecessary use of the genitive case		
Unreported error	Shuil Eoin isteach seomra folctha.			Maybe you should have **sa** after the word **isteach**

Table 2. Error Types and WCC Changes

Table 3 shows some sample learner text and some of the key error phrases used for spelling errors. The fact that neither *Gramadóir* nor the WCC was able to detect the word 'picture' is interesting, as it shows that they do not handle code-mixing, which would be quite common amongst primary school learners. This could be an area of future interest.

Error Phrases	Example	Source	*Gramadóir*	WCC
Do you mean	Nior tharraing sé	Learner	Do you mean /níor/ ?	**Níor**
Unknown word	Torraing	Learner	Unknown word	**???**
Not in database	Picture	Learner	Not in database but may be a compound /pic+túr/?	**???**

Table 3 Key Error Phrases for Spelling Errors

The overall logic for the WCC is shown in Figure 5.

Use the local error checking routines) Read and process learner text Depending on configuration options …. - If External error checking on … check for external errors - if local error check on … check for local errors Display user text with grammar and spelling messages (if any)

Figure 5. Overall Logic for the WCC

A sample of student text in the WCC is shown in Figure 6.

Figure 6. Sample of Student Text in WCC

4 Deployment and Evaluation

4.1 Deployment

The VCC and the WCC were used by primary schools students in two English-medium schools in Ireland. One of the schools was a mainstream, standard school (School 1) and the other was a school in a disadvantaged area (School 2). Ethical approval was applied for and obtained from the University's Research Ethics Committee and the parents. 3rd (age 8 – 9) and 4th class students (age 9 - 10) from School 1 used the VCC and 4th class students from School 2 used the WCC. The students used the resources over a period of several months on an ad-hoc basis.

4.2 Evaluation

Evaluation in the CALL field is complex. Quantitative and qualitative evaluation and formative and summative evaluation are all important. The VCC and WCC were evaluated using several different criteria. The aim of using various different evaluation criteria was to try to evaluate the Irish CALL resources from different perspectives. Chapelle's (2001) and Colpaert's (2004) CALL evaluation criteria were used to evaluate the VCC and WCC as CALL artefacts. The ICT4LT (2005) website which provides a CALL software evaluation checklist, was also used. The limitations of the evaluations include that some of it is based on self-reporting by young learners and that it was a small scale study with irregular and uneven usage.

The VCC was evaluated by the teacher and students in the mainstream school. An anonymous questionnaire-based survey was completed by 20 students (6 students were missing on the day of the questionnaire). There were both open and closed questions and students were encouraged to provide (negative) feedback. With regards to the VCC, the students 40% liked the tool, 45% liked it a little and only 15% did not like it. The majority found it helpful (45%) or a little helpful (35%), with only 20% saying it was not helpful. The majority preferred the animate mode (60%), over the static mode (15%), with 10% slightly preferring the animated mode, while 15% did not see the animated pages. The teacher found the resource useful as it was aligned with her teaching objectives for the class. Table 4 shows a summary of the student feedback on the VCC.

Question	Yes	No	A little/Both	Didn't see
Did you like the verb lessons?	40%	15%	45%	
Did you find them helpful?	45%	20%	35%	
Do you prefer the animated mode?	60%	15%	10%	15%

Table 4. Student Feedback on the Verb Conjugation Component

Students were also asked to give feedback after doing exercises/games with the VCC. The total number of students who answered online was 22 (note that not all students answered all the questions). Most students (84%) reported that they found the verb pages at least somewhat helpful, with little difference between those who viewed the pages in static and animated modes. It is interesting to note that more static mode students (26%) than animated mode students (10%) found the exercise/game hard. Table 5 shows a summary of the online student feedback data on the VCC.

Question	No	A Little	Yes
Did you find the verb lessons helpful?	16%	35%	49%
Static:	17%	35%	48%
Animated:		35%	50%
Did you like the end of lesson games?	12%	19%	70%
Static:	13%	13%	65%
Animated:	10%	25%	74%
Did you find the end of lesson games hard?	46%	36%	18%
Static:	35%	39%	26%
Animated:	57%	33%	10%

Table 5. Student Online Feedback on the Verb Conjugation Component

Students who did not find the VCC helpful said that they know the verbs already or that it was boring. Those who found it helpful said it "shows and tell what it means" and another reported that it cleared up confusion ("I was always getting confused and now I'm not"). When asked about their preference between static and animated mode, students who preferred static mode said that they understand it when the teacher explains it or that they found the animation mode annoying. Those who liked the animated mode said it was more enjoyable and it helped them. A summary of students' comments about the VCC are shown in Table 6. Note that the comments are provided as written by the students.

Did you find the Verb part helpful? **No:** Know already Too boring	**Did you find the Verb part helpful?** **Yes:** Shows and tells what it means Tells you how to spell them and more I was always getting confused and now I'm not
Which mode do you prefer? **Static:** I get it when the teacher tells me It's annoying	**Which mode do you prefer?** **Animated:** You would know more past tense verbs More fun Makes me understand It helps I kept on forgetting the h It will get you used to putting in silent letters
What was the best part and why? Games: learn stuff in games, fun	**What was the least enjoyable part and why?** Some games too part (paraphrased)

Table 6. Students' Comments on the Verb Conjugation Component

The teacher also provided an evaluation of the VCC. She said that it had sufficient learning potential because it focused on verb conjugation forms and her students did well in the verb exercises. She thought it was suitable for the learners, it was sufficiently challenging for them, it had the right level of difficulty and that the tasks were appropriate for them. The teacher said that explicit exposure to verb conjugation forms was pedagogically appropriate for her students. Note that another teacher was also involved in using and evaluating the VCC, but for external reasons was not able to use the resource to any great extent and the findings from her class are excluded from the evaluation.

Students in both schools used the WCC, but the findings here relate to the students in the mainstream school, as the numbers who used the WCC in the disadvantaged school were limited. The learners were asked to provide their feedback on the WCC via an anonymous open and closed questionnaire. Nineteen students completed the survey (7 students were absent on the day of the survey). Students reported that they liked using the WCC (yes (20%) and a little (50%), but 28% did not like it and a minority (28%) did not find it helpful. A sizeable minority reported that they did not understand the grammar error messages (42%) and spelling error messages (32%) and therefore, not surprisingly, many (grammar 47%, spelling 30%) said that they did not find them helpful. Most students said that they corrected their grammar errors (75%) and spelling errors (59%), although the empirical data does confirm this. It must be noted that only 11% said they liked writing in Irish and a majority (63%) said they would prefer to write in their copy than use the WCC. Table 7 provides a summary of the student feedback on the WCC.

Question	Yes	A Little	No
Did you like using the WCC?	22%	50%	28%
Did you find the WCC helpful?	44%	28%	28%
Did you understand the grammar error messages?	16%	42%	42%
Did you understand the spelling error messages?	26%	42%	32%
Did you find the grammar error messages helpful?	29%	24%	47%
Did you find the spelling error messages helpful?	35%	35%	30%
Did you correct your grammar errors?	75%		25%
Did you correct your spelling errors?	59%		41%
Do you like writing in Irish?	11%	47%	42%
Would you prefer to write in your copy?	63%		37%

Table 7. Student Feedback on the Writing Checker Component

Some of the reasons given for not finding it helpful included: “it was boring/hard”, “I already know how to write” or “I don’t like writing". Those who thought it was helpful said it told them the errors in their texts. Table 8 shows some of the students’ comments on the WCC. Note the comments are paraphrased, based on comments provided by the students.

Question	Finding
Why do you like/dislike writing in Irish?	Like: It’s our national language Dislike: Hard, boring, hard spellings, accents
Would you prefer the WCC or your copy for writing?	WCC: tells you your mistakes Copy: easier, faster, no keyboard problems

Table 8. Student Comments on the Writing Checker Component

The mainstream school teacher also completed a questionnaire and the feedback was positive. The teacher said that the WCC was beneficial for the students and enabled the students to construct sentences and stories. She felt that it was at an appropriate level for the learners as all the students could use the software. She said that it helped to consolidate classroom work. She said the main problem was that she did not know enough about computers herself. The teacher in the disadvantaged school initially came up with the idea to distinguish between grammar and spelling errors, as spelling

errors were not a priority for her. There were logistical difficulties for the teacher in that only four students (out of 17) were considered sufficiently competent to use and benefit from the WCC. Another difficulty was the fact that the school computer lab was closed during the project academic year and students had to travel to another venue to actually use the WCC – this obviously is not ideal.

Although both schools were boys-only schools in the same city, there are some significant differences between them. In the mainstream school, the students use the recommended textbook for their class, while in the disadvantaged school the students use a textbook for a more junior year. Also, more students are exempt from studying Irish in the disadvantaged school and there are fewer above-average students. Classroom management is more difficult and there are students leaving and returning to class from attending sessions with special needs teachers. This highlights the need to have flexible resources that can be used as the teacher sees fit. While the teacher in the disadvantaged school appreciated what the CALL resources can provide, their usage would probably be on a more ad-hoc basis than in the mainstream school.

From a CALL development point of view, it was relatively straightforward to use both Irish NLP tools. The IFSTME provides comprehensive information on Irish verbs. For pedagogical reasons, the VCC only uses a small subset of the information. The students were learning only a limited set of verbs, mainly regular verbs and some important irregular ones. In theory, the VCC could be modified easily to incorporate a more complete list of verbs, persons and tenses (although this was not required for this group of students). There were some difficulties in mapping and interpreting the conjugation changes for irregular verbs, but it must be noted that the IFSTME was not intended as a verb conjugation mechanism. It was used in white-box mode (i.e. some internal knowledge of the software was required for the VCC), but overall it was worthwhile using the IFSTME. Likewise, *Gramadóir* was a useful NLP resource for developing the WCC. It was robust and reliable and it would not have been possible to build the WCC without it.

5 Discussion

The VCC and the WCC demonstrate that it is possible and feasible to develop pedagogical, targeted NLP CALL resources for Irish. It helped that the two NLP tools used were robust and of a high quality. The learners and teachers were unaware of the underlying technology (and this is desirable). However, as is often the case, the problems were logistical rather than technical (Egbert et al., 2002; Ward, 2007). Access to computers and "space in the timetable" hindered the continued deployment of the Irish CALL resources.

It is important for NLP researchers working with any language to disseminate their findings and make their resources available to people outside the NLP community. It is even more important for NLP researchers working with minority languages to do so, as the resources are usually limited (Woodbury, 2003; Lam et al., 2014) and the pool of people working with the language small. Speakers, learners and other interested parties of minority languages are used to trying to do a lot with a little, and making NLP resources available to them could lead to the development of resources not initially envisioned by the NLP researchers.

6 Conclusion

This paper reports on how two NLP resources for Irish (i.e. the IFSTME and *Gramadóir*) were used to develop CALL resources for primary school children learning Irish. It shows that these NLP resources for Irish can be adapted and used to develop appropriate CALL resources. In order for the CALL materials to be successful, it is important that there is a seamless integration of the NLP tools in the CALL resources, so that the learner is unaware of their existence. Suitable, robust and accurate NLP resources are required, if the CALL materials are to work in a real deployment situation. The CALL resource should not fail or be inaccurate. The integration of the CALL resources with the curriculum itself is key if the resources are actually going to be used by the teacher and the students (Bull and Zakrzewski, 1997, Mc Carthy, 1999; Ward, 2007). This applies regardless of the language being studied – if the CALL resources do not help the teacher and aligned with the curriculum, they will not be used. There are other, non-technical, non-NLP related factors that help or hinder the actual usage of CALL resources. It should be noted that in order for the NLP resources to be used in the first place, there needs to be an awareness of their existence - teachers and CALL developers must know that

relevant NLP resources are available. This places an onus on NLP researchers to disseminate their research and tools to a wider audience than perhaps they would normally address. They could interact with the CALL community via CALL conferences and especially with ICALL (Intelligent-CALL) researchers via their Special Interest Groups (SIGs), conferences and workshops. This is particularly pertinent in the minority and endangered language context (e.g. Irish and other Celtic languages), where technical, financial and researcher resources are limited.

References

BBC. 2014. *BBC Languages (World Service English)*. Available at: http://www.bbc.co.uk/worldservice/learningenglish/

Joanna Bull and Stan Zakrzewski. 1997. Implementing learning technologies: a university-wide approach. *Active Learning*, *6*, 15-19.

Carol Chapelle. 2001. *Computer applications in second language acquisition: Foundations for teaching testing and research*. Cambridge: CUP

Jozef Colpaert. 2004. *Design of online interactive language courseware: conceptualization, specification and prototyping: research into the impact of linguistic-didactic functionality on software architecture*.- Antwerpen: Universiteit Antwerpen, Faculteit Letteren en Wijsbegeerte, Departement Taalkunde, 2004 - Doctoral dissertation

DCRGA. 2009. *20-year strategy for the Irish Language*. Prepared for the Department of Community, Rural and Gaeltacht Affairs. Fiontar, Dublin City University.

Joy Egbert,, Trena M. Paulus and Yoko Nakamichi. 2002. The impact of CALL instruction on classroom computer use: A foundation for rethinking technology in teacher education. *Language Learning & Technology*, *6*(3), 108-126.

John Harris and Lelia Murtagh. 1999. *Teaching and Learning Irish in Primary School*. Dublin: ITÉ.

Tina Hickey and Nancy Stenson.. 2011. Irish orthography: what do teachers and learners need to know about it, and why?. *Language, Culture and Curriculum*,*24*(1).

Thomas Koller. 2004. Creation and evaluation of animated grammars. Eurocall 2004, Vienna, Austria (3rd September 2004).

ICT4LT. 2005. *Information and Communications Technology for Language Teaching (ICT4LT) Project: Evaluation Forms*

Lori S. Levin and David A. Evans. 1995. ALICE-cha: A Case Study in ICALL Theory and Practice. *In*: V. M. Holland, J.D. Kaplan and M.R. Sams (Eds.) *Intelligent Language Tutors* (pp. 327-44). Mahwah: Lawrence Erlbaum

Khang N. Lam, Feras Al Tarouti and Jugal Kalita. 2014. Creating Lexical Resources for Endangered Languages. *ComputEL at ACL 2014*.

Claudia Leacock., Martin Chodorow, Michael Gamon and Joel Tetreault. 2014. Automated grammatical error detection for language learners. *Synthesis lectures on human language technologies.* Second Edition.

C. Matthews. 1993. Grammar frameworks in Intelligent CALL. *CALICO Journal* 11, 1: 5-27.

Brian McCarthy. 1999. Integration: the sine qua non of CALL. *CALL-EJ Online*,*1*(2), 1-12.

Merriampark. 2005. *Levenshtein distance algorithm*.

Linda Murphy and Stella Hurd. 2011. Fostering learner autonomy and motivation in blended teaching. In: Nicolson, Margaret; Murphy, Linda and Southgate, Margaret eds. *Language Teaching in Blended Contexts.*Edinburgh, U.K.: Dunedin Academic Press Ltd, pp. 43–56.

Lelia Murtagh. 2003. *Retention and Attrition of Irish as a Second Language: a longitudinal study of general and communicative proficiency in Irish among second level school leavers and the influence of instructional background, language use and attitude/motivation variables*. PhD thesis, University of Groningen.

John Nerbonne. 2003. Natural Language Processing in Computer-Aided Language Learning. *In:* S. Mitkov (Ed.) *The Oxford Handbook of Computational Linguistics*. Oxford: OUP

Donncha Ó hÉallaithe. 2004. From Language Revival to Survival. In: C. M. Murchaidh (ed.), *Who Needs Irish? Reflections on the Importance of the Irish Language Today*, Dublin: Veritas Publications.

Pádraig Ó Riagáin. and M. Ó Gliasáin. 1994. National Survey on Languages 1993: Preliminary Report. *Dublin: Institiuid Teangeolaiochta Eireann*.

Kevin Scannell. 2005. *An Gramadóir*. Available at: http://borel.slu.edu/gramadoir/

Thornbury H., Elder M., Crowe D., Bennett P. & Belton V. 1996. Suggestions for successful integration. *Active Learning*, 4, 18-23.

Elaine Uí Dhonnchadha. 2002. Two-level Finite-State Morphology for Irish, *In: Proceedings of LREC 2002 3 rd International Conference on Language resources and Evaluation*. Las Palmas de Gran Canaria, Spain.

Anne Vandeventer Faltin. 2003. *Syntactic Error Diagnosis in the context of Computer Assisted Language Learning*. PhD Thesis, Faculté des letters d l'Université de Genève.

Monica Ward. 2001. *A Template for CALL Programs for Endangered Languages*. Masters thesis Dublin City University.

Monica Ward. 2007. *A template for CALL programs for endangered languages.* (Doctoral dissertation, Dublin City University).

Tony Woodbury. 2003. Defining documentary linguistics. *Language documentation and description*, *1*(1), 35-51.