How NLP Can Strengthen Digital Game-Based Language Learning Resources for Less Resourced Languages

Monica Ward[1], Liang Xu[1], Elaine Uí Dhonnchadha[2]
[1] Dublin City University, Ireland
[2] Trinity College Dublin, Ireland
monica.ward@dcu.ie, liang.xu6@mail.dcu.ie, uidhonne@tcd.ie

Abstract

This paper provides an overview of the Cipher engine which enables the development of a Digital Educational Game (DEG) based on noticing ciphers or patterns in texts. The Cipher engine was used to develop the *Cipher: Faoi Gheasa*, a digital educational game for Irish, which incorporates NLP resources and is informed by Digital Game-Based Language Learning (DGBLL) and Computer-Assisted Language Learning (CALL) research. The paper outlines six phases where NLP has strengthened the *Cipher: Faoi Gheasa* game. It shows how the Cipher engine can be used to build a Cipher game for other languages, particularly low-resourced and endangered languages in which NLP resources are under-developed or few in number.

Keywords: Digital educational game, digital game-based language learning, low-resourced language, computer-assisted language learning

1. Introduction

NLP technologies are designed to interact with human language for a variety of purposes including speech recognition, natural language understanding and natural language generation. Digital Educational Games (DEG) involve the use of computer game characteristics for educational purposes with a focus on learner engagement. NLP technologies can be used in games to improve the experience for players, especially for Digital Game-Based Language Learning (DGBLL) in which learning languages is a desired side effect of playing games. Computer-Assisted Language Learning (CALL) involves the use of technology in the language learning process. It encompasses DGBLL but is broader in reach as it also covers other approaches to language learning. CALL resources can be used in traditional learning settings as well as online, offline and in individual learning spaces. In the paper, we discuss how players can be encouraged to notice errors (ciphers) in texts and in a game. This allows them to become aware of errors in Irish spelling and grammar which is something they often do not notice or are unaware of. This game can provide valuable learning opportunities for language learners including vocabulary, reading and writing practice. Section 2 provides an overview of the role of motivation in language learning and particularly the issue of reluctant learners of Irish. It looks at Digital Game-Based Language Learning and NLP-infused games in CALL. Section 3 describes the methodology and gives a brief explanation of the Cipher engine and the six phases in the Cipher pipeline with an NLP component. Section 4 outlines the implementation of the game, while Section 5 provides the results and evaluation of the game. Section 6 covers the discussion and section 7 is future work. Section 8 gives the overall conclusion of the paper.

2. Background and Related Work

2.1 Motivation

Language learning can be interesting, stimulating and fun. However, for many learners who have to learn a language it can be uncomfortable, boring and not enjoyable. Many second language (L2) learners are reluctant learners particularly when it is a compulsory subject in their education system. Some of the world's L2 English learners fall into this category. There are several challenges for these reluctant learners. Protheroe (2004) notes that they do not complete tasks and they avoid challenges, even though they are capable of excelling. Sanacore (2007) highlights the importance of fostering intrinsic motivation with reluctant learners. He makes four suggestions to help reluctant learners. He states that it is important to create a learning environment that is encouraging and challenging. Sometimes teachers think they have to simplify things to the point at which there is no challenge for the student, but even beginning language learners should have opportunities to engage in challenging learning activities. Students should be provided with opportunities to make learning choices. Student choice brings many positive benefits including increasing their autonomous behaviour, eliciting cognitive flexibility, high task interest, positive emotion, creativity, and persistence (Clifford 2007; Deci and Ryan 1987). Sometimes students are passive learners in their lessons and Sanacore (2007) advocates for increasing students' participation in classroom activities as a way to foster intrinsic motivation. The aim of a teacher should be to encourage students to love learning as this will make learning activities more enjoyable and fruitful for students.

Although Irish is one of the three official languages of Ireland (English and Irish Sign Language are the other two), it is only spoken by 1.5% of the population outside of the education system on a daily basis (CSO, 2016). With some exceptions, it is a compulsory subject in both primary and post-primary schools. Students often lack motivation to study the language and this can make Irish lessons seem like drudgery rather than an enjoyable experience for some students. The teachers, who in the majority of cases are not native speakers, often have to shoulder the burden of Irish language education as many parents are not able to help their children with their Irish homework. Until recently, there were very few interactive resources for teaching Irish in the school context. There are now some online resources linked to specific textbooks, but these are limited in number. Irish is not considered as a Modern Foreign Language (MFL) in the Irish school system which means that some of the pedagogical innovations from the MFL field do not find their way into the teaching of Irish. This is unfortunate, as many of these innovations can help to make the learning experience more engaging and personalised for the students (e.g., Ward et al., 2019). In a

recent report, Irish language inspectors noted the need for enhanced Irish language learning activities.

> "*Inspectors observed an overreliance on translation from Irish to English as part of the pupils' experience of Irish and highlighted a need for more fun and engaging Irish language learning activities.*" - The Irish Times (23/03/2022)

In the Irish context, just as Sanacore (2007) outlined for reluctant learners in general, there is a need to provide challenging learning activities, offer student choice and provide opportunities for more active learning. One way of providing these activities and opportunities is with the use of a Digital Game-Based Language Learning (DGBLL) app for Irish.

2.2 Digital Game-Based Language Learning

Sørensen and Meyer (2007) discussed the trend of technology in language learning moving away from rote-based acquisition that focuses on drills, grammatical structures and translation tests to context-based acquisition that focuses on task-based, project-based and content-based approaches. Games can be seen as a lever for the transformation. In fact, game progression depends on players' performance of skills which are based on their actions in games rather than simply memorising information or giving correct answers (Dunkel, 1991). Performance is expected in game-based activities while schools tend to focus on tests and competence (Gee, 2005). For instance, people may be more willing to read game-related text (e.g., in-game text, game walkthrough, supplementary materials) in order to win a game instead of reading a linear text assigned to them for a test and in this way their comprehension would increase with each repetition (Underwood, 1987).

The idea of digital game-based learning (DGBLL) has proven to be supportive for L2 learning. In recent years, an increasing number of studies have analysed the impact of digital games on L2 developments and these studies have led to several meta-analysis studies. Findings from these studies have reached the conclusion that digital games have positive effects on L2 learning, particularly on L2 vocabulary (Dixon et al., 2022; Chen et al., 2018; Tsai and Tsai, 2018). Dixon et al., (2022) suggested that DGBLL works better for games designed for entertainment than games designed for educational purposes and that the latter has been overlooked by the CALL community. This is mainly due to the fact that DEGs lack commercial interest from industry and the too "obvious" objectives of language learning further limits a DEG's success (Reinhardt, 2019). Moreover, DEGs are often overshadowed by games designed for entertainment when it comes to engaging elements (e.g., storylines) and authentic language interaction (e.g., spoken and written input) (Dixon et al., 2022).

However, DEGs have the advantage of providing learners with access to minority languages. The language options of games designed for entertainment are very limited and are mainly available for 'bigger' languages e.g., English and Spanish. For games designed for entertainment, the incentive for adding other languages would be much lower compared to DEGs or entertainment apps specifically designed for language learning purposes. For example, there are more than 40 different languages in Duolingo including many minority languages according to the 2021 Duolingo language report (Blanco, 2021) and the number has been increasing. Therefore, DEGs can be important for supporting the learning and teaching of minority languages.

2.3 NLP Infused Games and CALL

NLP resources have the potential to contribute to Computer-Assisted Language Learning (CALL) (Ward, 2019), but this potential remains largely under-utilised. While some NLP resources have been used in CALL to date, this is often as a by-product of NLP research rather than an explicitly stated aim of the NLP resources. A basic impediment to the use of NLP resources in CALL is that there is limited overlap and interaction between the NLP and CALL communities in general. NLP researchers focus mainly on language and linguistic technologies (with limited consideration for pedagogy) while CALL researchers prefer to focus on pedagogy first and technology second.

There are many reasons for the limited use of NLP resources in CALL, but the two overarching ones are that the development of CALL resources is challenging (Godwin-Jones, 2015) and the overall difficulty of incorporating NLP technologies into CALL resources (Heift & Schulze, 2007). Ideally, the development of CALL resources would involve a range of experts including language teachers, linguists, pedagogical designers, software engineers, user interface (UI) designers and educational technologists amongst others. It is rare for CALL teams to have access to this range of experts, particularly given the limited time and financial resources available to most CALL development teams. The development of NLP resources is also challenging and most NLP resources are designed to deal with native or near-native speakers rather than foreign language learners. The errors that are inherent in language produced by learners can be very challenging for NLP resources that expect native-like language input. Language learning involves five main language skills (reading, writing, listening, speaking and interacting) and NLP has the potential to contribute to CALL in all of these areas (Ward, 2017; Ward, 2019). Many language learners currently use NLP-embedded tools for writing (e.g., spelling and grammar checkers (Ferris et al., 2013)) and text-to-speech tools can also be helpful to learners (e.g., Cardoso et al., 2015), especially if they are unfamiliar with the L2 writing system.

3. Methodology

3.1 Cipher Engine

CALL research is heavily focussed on the most commonly taught languages, particularly English. This is not surprising as there are around 1.5 billion English language learners in the world (British Council, 2014). This means that most of the CALL resources developed are for learners of English and to a lesser extent Spanish, French and German. Consequently, there are fewer resources for Less Commonly Taught Languages (LCTLs) (Ward, 2015), which can range from languages with a large number of speakers such as Chinese and Arabic to languages widely spoken in their country of origin such as Polish and Thai. The term LCTL also covers minority or regional languages like Catalan and endangered languages like North Saami

(see Ward, 2018 for more details). A language can be an official language of a country and yet be a Less Commonly Taught language. This is the case for Irish. Notwithstanding the large Irish diaspora, on a worldwide scale there are very few learners of Irish. It is more challenging to develop CALL resources for LCTLs (Ward, 2015) and therefore LCLT CALL researchers sometimes are creative (e.g., Millour et al., 2020) or aim to leverage existing resources and adapt them to their own LCTL (e.g., Purgina et al., 2017).

This is the case in the development of *Cipher: Faoi Gheasa*, which was based on the original Cipher game (Xu and Chamberlain, 2020) for advanced level English language users (B1 - C2, on the Council of Europe CEFR scale). Cipher is a crowdsourcing game designed for identifying errors in text which uses the idea of 'games with a purpose' (GWAP) (Von Ahn, 2006). Error spotting was gamified such that people were encouraged to spot errors in texts through the game. While playing the game, players are making annotations to the text, and thus data is collected for further analysis. The results showed that people could easily notice text errors in the game and it is therefore possible to identify errors using a game. Moreover, feedback from users indicated that Cipher was fun to play and has potential to help language learning.

Cipher: Faoi Gheasa was adapted to cater for Irish language learners of A2-B1 level. A new storyline, new game features and elements, and updated rules were added to the original Cipher engine to encourage language learning and facilitate in-game data collection. The theme of "reconnecting to the spirit of the language" (Napier & Whiskeyjack, 2021) functions as the socio-cultural background behind the game design.

3.2 A Language Independent Cipher Engine

It can be more challenging to develop DGBLL resources for Less Commonly Taught Languages (LCTLs) and therefore where possible, DGBLL developers should aim to develop resources that are language independent. In other words, the framework should be decoupled from the language so that a plug and play approach can be used. With this approach, language specific modules can be added to the template to create a DGBL resource for that specific language. While pedagogical issues may arise due to the range of human languages, where possible a language independent approach is beneficial for the development of DGBLL resources for LCTLs and this is the approach adopted in the design and development of the Cipher engine.

3.3 Game Mechanics

We focus on three language tasks in this game: noticing, reading, and writing. Integrated with interesting game elements, these language tasks are mapped onto game tasks which fit into the game storyline and the theme of "reconnecting to the spirit of the language". The storyline is as follows:

1. There is an **evil game character** whose goal is to make ancient tales unreadable to people by casting spells upon the tales in which many ancient mythological creatures dwell. The evil spirit, Syfer, wishes to make people forget the tales and ensure that these mythical beings will eventually vanish as their existence is based on people's belief in them. The aim of the player is to defeat the evil Syfer.

2. The players need to **read stories** and **find the enchanted words** and **identify the spells** that were cast upon the words.

3. The design of the spells is inspired by the idea of **steganography** following the original Cipher game (Xu and Chamberlain, 2020). A spell changes certain words (these modified words are known as enchanted words in the game) in the story in a particular way so that the players can identify a spell by finding patterns of errors in the story. This can help the practice of spelling and reading. For ease of understanding, we also refer to spells as ciphers in this paper.

4. Some spells (ciphers) are associated with the **grammatical information** (e.g., word gender), which is designed to help learners get to know more about the vocabulary.

5. If the players fail to find a spell, they will be asked to change the ending of the story in order to delay forgetting of the story and of the magical beings involved. This is designed to help the practice of **writing**.

In addition, there are *power-ups* available for the players to use in case they get stuck. Besides adding more fun to the game, the design of power-ups enables players with little or no knowledge of the language to still be able to enjoy the game. In summary, the task of the player in the game is to find the enchanted words and identify spells cast by Syfer (the evil game character) and thereby save the ancient spirits and the stories. The incentive of the game is to gain spirit power which is a token in the game. Players will gain tokens if they do the "right things" in the game, which includes finding errors, finding ciphers and continuing the story. Players' tokens will be deducted by clicking a word that is not an error, clicking a cipher that is not "responsible" for the errors, using power-ups or abandoning a story.

Cipher: Faoi Gheasa adapts to the player's language level. If a text is too easy for a player, they are shown a more difficult text the next time. Conversely, if a text is too difficult for a player, they are shown an easier text the next time. Furthermore, the difficulties of ciphers are adaptive. The choice of ciphers used in the game text is reflected by players' performance. Figure 1 shows the logic of the adaptability of *Cipher: Faoi Gheasa.*
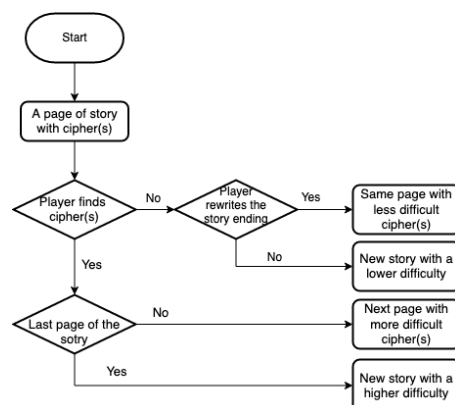

Figure 1: Adaptivity in *Cipher: Faoi Gheasa* using level analyser

4. Implementation

NLP resources and tools have contributed in different ways to the development of *Cipher: Faoi Gheasa*. For example, POS tagging is used for text level analysis, cipher detection analysis, and the analysis of player written texts in learner corpus collection.

4.1 Pre-Processing

It is important to ensure that the texts presented to the players have no spelling or grammatical errors (apart from those introduced by the Cipher engine). There are currently three main sources of texts for *Cipher: Faoi Gheasa*: Dúchas[1] texts, original texts based on traditional Irish stories and international fairy tales and a limited number of texts from the Gutenberg Project[2]. The selected Dúchas texts were stories written by children in the 1940s. These texts were not written in the Official Standard for Irish (*An Caighdeán Oifigiúil[3])* and needed to be converted to the modern standard. NLP tools were used to detect spelling combinations that needed to be updated. In addition, some of the older texts used the Gaelic font, for example, ċ needed to be replaced by ch. Once these changes had been made, all the texts were reviewed for errors using the online the electronic version of *Ó Dónaill's Irish-English Dictionary[4]* and *Gramadóir[5]* spelling and grammar checker for Irish. Figure 2 shows this preprocessing step.
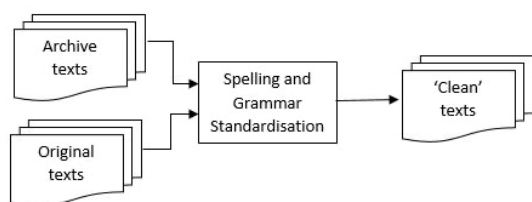

Figure 2: Pre-Processing step

4.2 POS tagging

A POS tagger for Irish (Uí Dhonnchadha and van Genabith, 2006) provides XML formatted POS tagged text to the Cipher engine so it can choose to highlight particular parts of speech. The tagger which was initially developed for general Irish texts, can provide useful information for educational purposes as well. Figure 3 shows the POS-tagging step.

Noun gender is important when learning Irish vocabulary, but it is rarely taught explicitly in schools and students are often unaware of the concept of gender in Irish. Most of the Irish language learners in Ireland are L1 English speakers and they are unfamiliar with the concept of grammatical gender. In *Cipher: Faoi Gheasa*, we draw attention to the gender of nouns by highlighting masculine and feminine nouns in different colours.
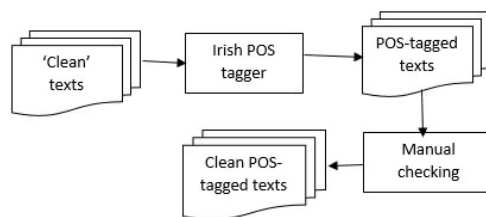

Figure 3: POS-tagging step

Figure 4 shows an example where the word *dullahan* 'dark man', a masculine noun is highlighted in sapphire blue (representing the colour of the Water Spirit) and the word *murúch* 'mermaid', a feminine noun is highlighted in ruby red (representing the colour of the Fire Spirit). Figure 4 explains gender highlighting in the context of the game in a way that fits into the game storyline.

Figure 4: Game storyline explanation of noun gender highlighting

Figure 5 shows a page of a story where all masculine nouns (e.g., *lá* 'day' and *ocras* 'hunger') are coloured blue while the feminine nouns (*tine*, 'fire' and *cailleach* 'witch') are coloured red.

Figure 5: Cipher screen with ciphers and gender highlighted

The names Hansel and Gretel are also coloured according to their gender but words tagged as proper nouns by the POS tagger are excluded from ciphers as there is variability in the ways in which names are spelled and it can be difficult to determine whether a cipher has been applied or not. For illustration purposes the ciphers and their correct forms are shown in Figure 5.

Normally, when a player is playing the game, the correct forms are not shown (see Figure 6) unless power-ups (see Figure 7) are used.

Figure 6: Cipher screen as seen by players

Figure 7: A screenshot of power-ups

There is also a version of *Cipher: Faoi Gheasa* with an Irish language interface but the English language version is shown here for illustration purposes. The Cipher engine can be easily reconfigured to focus on different aspects of language as desired e.g., noun plurals or particular verb tenses.

4.3 Text level analyser

Vygotsky's (1978) Zone of Proximal Development (ZPD) is an important concept in learning in general and is very relevant in language learning contexts. In the Cipher game, it is important that learners are presented with texts at the appropriate level for their language ability. If a text is too hard, the learners will be demotivated and will not want to play the game. If it is too easy, they will be disinterested. A text that has a level of linguistic difficulty that is suitable for the learner will be most engaging for them and will incentivise them to play the game. A combination of NLP tools provides information that can be used to determine the linguistic complexity of a piece of text. There are several checkers available for this in English e.g., Flesch–Kincaid readability tests (Kincaid et al, 1975). There are currently no publicly available text analysis tools for Irish, however the Irish NLP tools[6] are used to provide information about lexical and grammatical complexity which is used to rank the Irish texts used in *Cipher: Faoi Gheasa*. Figure 8 shows the steps in the lexical analysis process.
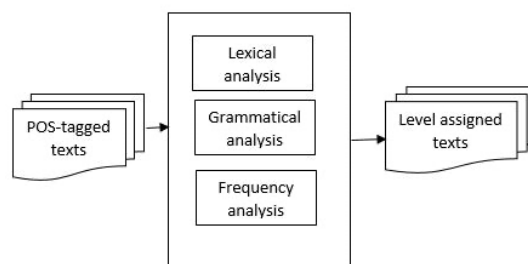

Figure 8: Level analysis phase

4.4 Analysis of Player Cipher Detection

Once the players have played *Cipher: Faoi Gheasa*, their game metrics are stored for analysis. It is helpful to use the Irish POS tagger to classify this data to get better insights into the level of knowledge and awareness that the players have of Irish spelling. Figure 9 shows the steps in the analysis of the cipher detection phase. There are three specific metrics that are calculated:

1. Ciphers correctly identified by players (true positives)
2. Ciphers missed by players (false negatives)
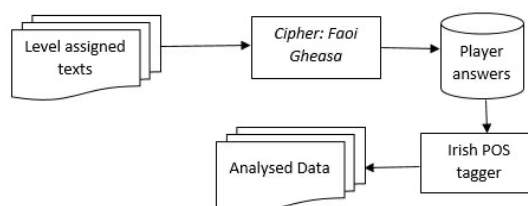3. Ciphers incorrectly identified by players (false positives)

Figure 9: Analysis of cipher detection phase

4.5 Analysis of Player Texts

If players fail to identify a sufficient number of ciphers in a page of text, they have the option of entering a sentence to change the ending of the story. The sentences entered in this way can be analysed to provide further insight into the players' understanding of the text they have read and to give some insight into their level of Irish. The Irish POS tagger can detect if the text is correct Irish (with POS information), incorrect Irish or English (see Section 5.4 for details). Figure 10 shows the analysis phase of players' texts.
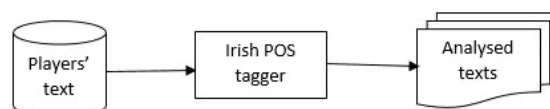

Figure 10: Analysis of players' texts

4.6 Learner Corpus Collection

The sentences entered by the players can be collated to form a corpus of learner Irish. Currently, there is no such publicly available learner corpus from a game for Irish. The

---

[6] https://www.scss.tcd.ie/~uidhonne/irish.utf8.htm

44

use of *Cipher: Faoi Gheasa* will enable the development of such a corpus.

4.7 NLP Pipeline

The NLP pipeline can help to build other versions of Cipher that are language-specific and culture-specific. The game will work the same way but can be customised for different languages. Figure 11 provides an overview of the NLP pipeline for *Cipher: Faoi Gheasa*. It shows the role of each NLP component in the creation of *Cipher: Faoi Gheasa* and the subsequent analysis of the players' actions while playing the game.
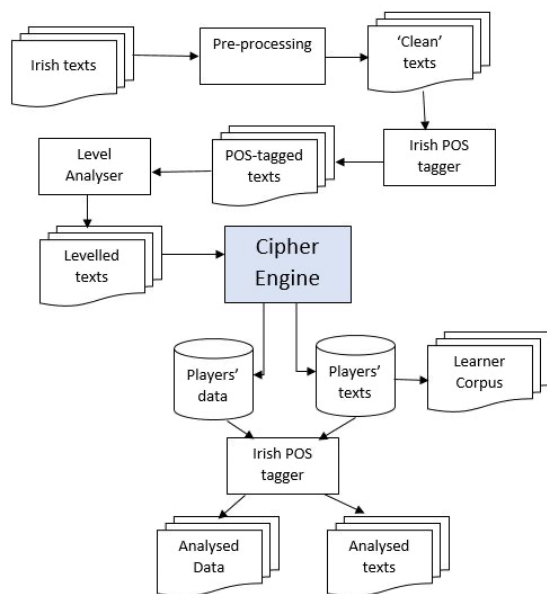

Figure 11: NLP pipeline for Cipher: Faoi: Gheasa

4.8 Choice of materials

We chose the theme of stories and myths for this game as we wish to engage the learners and hold their attention. By choosing familiar fairy tales at the lower levels we can build the learner's knowledge from their L1. At the more advanced levels we use folk tales and mythology which are engaging and can be made culture-specific and reflects the theme of "reconnecting to the spirit of the language". Currently we are using stories from two online archives: www.Dúchas.ie and www.Gutenberg.org, as well as some Irish versions of well-known fairy tales.

5. Results and Evaluation

5.1 POS tagging

The Cipher engine was able to use the XML-formatted POS tagged texts directly and could generate the relevant highlighting features without difficulty.

5.2 Text Level Analyser

This information was useful in deciding which texts to show players. It was particularly helpful for the beginner level as sometimes it can be difficult to realise how limited beginner level students' language knowledge actually is. It can be tempting to add elaborate texts to the game, but if they are beyond the player's Zone of Proximal Development, then they will be off-putting for students. One student commented that "the level of irish was about

right but a few verbs that we didn't learn yet". In a pilot study in one primary school where nine classes of 10–12-year-olds tried out the game, approximately 47% of students who filled out a questionnaire think the difficulty level of the game text is about right. (Figure 12)
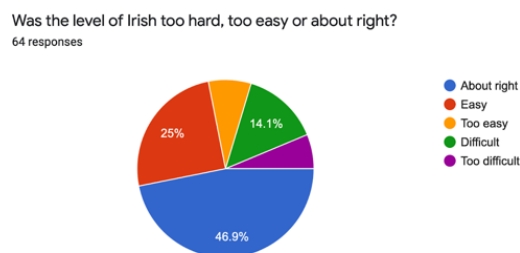

Figure 12: Students' opinions of the difficulty of the game text

5.3 Analysis of student cipher detection

5.3.1 Ciphers correctly identified by players

The POS tagger was useful in classifying the information on cipher detection by the players. In the case of occurrences of an error (where it was seen more than 20 times), there is a mix of POS categories in the correctly identified errors. Most of these words would be familiar to players (e.g., *choill* 'wood', *bhia* 'food' and *mhuc* 'pig'), while the ciphered words are unlikely (*scrao, uca*) or impossible in Irish (*hcaet, htiam*). Table 1 shows the top 10 ciphered words correctly identified by players.

Ciphered Word	Correct word	Lemma	Meaning	POS
lhoilc	choill	coill	forest	N(m)
scrao	ocras	ocras	hunger	N(m)
uca	acu	ag	at	Prep.
hcaet	teach	teach	house	N(m)
arsaa	arsa	arsa	said	V(past)
bhíí	bhí	bí	was	V(past)
niáeslim	milseán	milseán	sweet	N(m)
gaeb	beag	beag	small	Adj.
htiam	maith	maith	good	Adj.
hraibo	oraibh	ar	on you (pl)	Prep. Pron.

Table 1: Ciphers correctly identified by players

5.3.2 Ciphers Not Identified

There is also a mix of POS categories where the ciphers were not identified by the players. It is interesting to note that the words (*ocras* 'hunger' and *siad* 'said') are repeated in the lists, with different ciphers. Table 2 shows the top 6 ciphers missed by players.

Ciphered Word	Correct word	Lemma	Meaning	POS
ann	an	an	the	Art(d)
asra	arsa	arsa	said	V(past)
dias	siad	siad	them	Pro(3P)
dais	siad	siad	them	Pro(3P)
sarco	ocras	ocras	hunger	N(m)
nna	ann	ann	there	Adv.

Table 2: Ciphers not identified by players

5.3.3 Ciphers Identified Incorrectly

Table 3 shows the top 10 words incorrectly identified by players as ciphers, when they were in fact correctly spelt. The word most often incorrectly identified as an incorrect spelling was *arsa* 'said'. While this is commonly used in stories to indicate direct speech, it may not have featured very often in the students' textbooks. It is interesting to note that nouns (*choill*, *bhia*, *lapadaíl*) were frequently incorrectly identified as being ciphers, followed by adverbs (*anall*, *anonn*). Both *choill* and *bhia* have initial mutations and students may be more familiar with the unmutated lemma forms *coill* and *bia*.

Incorrect word	Lemma	Meaning	POS
arsa	arsa	said	V(PI)
clábar	clábar	mud	N(m)
choill	coill	forest	N(f)
bhia	bia	food	N(m)
lapadaíl	lapadaíl	paddling	N(m)
ó	ó	from	Prep.
rolladh	rolladh	roll	N(m)
anall	anall	back	Adv.
anonn	anonn	over	Adv.
mhuc	muc	pig	N(f)

Table 3: Top 10 words incorrectly selected as ciphers

5.4 Analysis of student texts

There were 184 sentences entered by the players. In order to encourage players to enter text and to avoid frustration at spelling errors in their own texts, a spelling checker feature was removed from this part of *Cipher: Faoi Gheasa*. This meant that the students could enter text of any kind - correct Irish, incorrect Irish, text in English or even nonsense text. An initial analysis of the players' texts indicates that there was actually quite a low percentage of texts (16%) that were in Irish and relevant to the story. The

most common type of text was in Irish but not relevant to the story (39%), while 18% was Irish junk text. Texts in English accounted for 16% of the texts entered while nonsense junk comprised 14% of the texts. Table 4 shows the breakdown of texts by category.

Text Type	%
Irish - not relevant to the story	39%
Irish - nonsense	18%
Irish - relevant to the story	16%
Nonsense	14%
English	13%

Table 4: Analysis of students' texts by type

5.5 Corpus

The players of *Cipher: Faoi Gheasa* are contributing to a corpus of Irish learner texts. This corpus will continue to grow as more users play the game. While the corpus is currently small, it does provide a starting point for future research and will provide insights for Irish language educators, particularly primary school teachers and teacher educators. Millour and Fort (2020) report on interesting approaches for a crowd sourcing approach for low resource language and there is potential to leverage some of their findings in future work in this area.

5.6 User Experience Study

The evaluation and user study were analysed from the aspect of gaming experience, learning experience, adaptivity and usability according to the four-dimension evaluation framework (Law and Sun, 2012). In a survey of primary school students who played the game (n=64), 62.5% of the participants were positive about learning Irish through the game, 57.8% said the game was easy to play and 59.4 percent of the participants were willing to read the stories in the game. The full details of the evaluation process and results can be found in the study (Xu et al., 2022), which focuses on the user experience study of *Cipher: Faoi Gheasa*.

6. Discussion

6.1 NLP Perspective

It is important to ensure the quality of the texts used in *Cipher: Faoi Gheasa*. It was helpful to have the Irish POS tagger (Uí Dhonnchadha and van Genabith, 2006) for tagging the Irish texts and analysing the players' Irish sentences, as well as for analysing the level of texts. *Gramadóir* (Scannell, 2007) is a useful tool for checking Irish texts for spelling and grammar errors. One of the motivations for using the game with students was to gain extra insights into their knowledge of Irish spellings and spelling errors. The Irish POS tagger was useful for identifying the POS categories of the ciphers detected, not detected and incorrectly detected by the players.

Many low-resourced languages will not have a POS tagger available to them. While it was helpful to have a tagger, for other less resourced languages, a linguist could manually

tag specific texts with the relevant XML tags which could then be fed to the Cipher engine.

6.2 Student Feedback

Feedback from students on the use of *Cipher: Faoi Gheasa* was positive. Given that almost all of their learning of Irish comes from classroom activities and printed textbooks, it was not surprising that they enjoy playing the app. Positive comments included *"its better than learning in the classroom"*. Asked what they liked about the game, one student replied *"not having to learn irish out of books"*.

When developing a GWAP, it is important that the game dynamics work for the players. Students understood the cipher storyline and context (*"the storyline is great"*).

Some sample comments from students indicate that they really enjoyed the game.

> *"this game is very good and fun it is also very very entertaining we would rather do irish on this app than from [name] book thank you very much"*

> *"i think its a fun game and i would like if we could play it school. It is very adventurous."*

> *"the joy in winning the astonishing game"*

Learner autonomy is a feature of *Cipher: Faoi Gheasa* and students can play at their own pace. More advanced students will move through the game quicker while other students can move at a slower pace. Learner autonomy is advocated by Sanacore (2007) as a way of motivating reluctant learners and it is interesting to see that some players themselves were able to articulate this: "*the fredom and i prefer to play games than just get told things*".

While players could just scan the texts looking for ciphers, based on some student comments, it is interesting to see that some students did read the texts and understand the stories. One player commented that "*the witch died, which is what we wanted to happen*".

Students studying Irish have very limited exposure to Irish outside of the classroom (Harris et al 2006, cited in Hickey & Stenson, 2011). While Irish reading is not the sole focus of *Cipher: Faoi Gheasa*, it does provide a novel and interesting way for the students to read Irish texts. Students would generally only see Irish in a textbook, which can be a bit staid for some learners. The digital format particularly appeals to some students.

Another feature of *Cipher: Faoi Gheasa* is that students can write sentences in Irish as part of the game. They would generally not write texts in a digital format in Irish so this is a novel feature for them. One additional point to note about the use of *Cipher: Faoi Gheasa* in the classroom context is that even students who are exempt from studying Irish showed an interest in the game. This is particularly satisfying as often these students can feel excluded from class when the teacher is teaching Irish as they are assigned other tasks to do instead of Irish. This is a positive unintended consequence of *Cipher: Faoi Gheasa* - a more inclusive approach to teaching Irish.

7. Future Work

The NLP aspects of *Cipher: Faoi Gheasa* worked well but there is room for improvement in terms of some of the game dynamics. One student commented that they were not able to save their progress (*"no way to save your progress"*). However, there is a way to save progress and this will be made more obvious to players in future. Some students wanted extra pizzazz in the game ("no *cool celebration*"). There were also issues to do with wifi connections and slightly old laptops which are obviously outside of the control of the developers but are issues that cannot be ignored nonetheless.

> *good game,but need more up to date laptop*

> *The game was kinda boring the hints dident help .... but it is very fun to play in school with your friends but it was ok nice game rate it a 5*

There is also a need to test *Cipher: Faoi Gheasa* with different types of schools. The players reported on in this paper were all from an English medium primary school (which make up the vast majority of Irish schools). It will be interesting to test *Cipher: Faoi Gheasa* with students from Irish-medium schools and also with students in Irish speaking regions of Ireland. *Cipher: Faoi Gheasa* could be suitable for adult learners as well and it will be tested with this cohort as well. We also intend to provide more texts in *Cipher: Faoi Gheasa* and to add new ciphers to the game. Also, it would be good to adapt *Cipher: Faoi Gheasa* to cater for the needs of A1 (complete beginner) students.

8. Conclusion

The development of Cipher: Faoi Gheasa was greatly facilitated by the use of the Cipher engine. The use of NLP tools and resources strengthened the game as they provided relevant information on parts of speech and enabled texts to be classified into suitable levels for learners. They also helped to ensure the quality of the texts presented to the players by identifying incorrect spellings in the texts at the preprocessing stages before they were provided to the Cipher engine. Students who have played the game reported that they enjoyed it and would like to continue to play it. This is encouraging as usually students will try to minimise the time they spend in class learning Irish. While developing NLP enhanced DGBLL apps for Less Resourced languages is more challenging, it is not impossible. This paper demonstrates that a structured and creative use of existing Irish NLP resources and generic NLP tools can be used to good effect to develop games that are pedagogically suitable and appropriate for language learners.

9. Acknowledgements

10. Bibliographical References

Blanco, C. (2021, December 13). The 2021 Duolingo Language Report. Duolingo Blog. Retrieved April 14, 2022, from https://blog.duolingo.com/2021-duolingo-language-report/

Breen, M. (1987). Leaner contribution to task design. In. C. Candlin& D Murphy. Language Learning Tasks. Englewood Cliffs, NJ: Pearson Prentice Hall.

British Council. (2014) English in Numbers. English in numbers | British Council. (n.d.). Retrieved April 14, 2022, from https://www.britishcouncil.cn/en/EnglishGreat/numbers

Central Statistics Office [CSO]. (2016). Census of population 2016–profile 10 education, skills and the Irish Language.

Chen, M. H., Tseng, W. T., & Hsiao, T. Y. (2018). The effectiveness of digital game-based vocabulary learning: A framework-based view of meta-analysis. British Journal of Educational Technology, 49(1), 69-77.

Council of Europe. (2022). The CEFR Levels. Common European Framework of Reference for Languages (CEFR). (n.d.). Retrieved April 14, 2022, from https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions

Dixon, D. H., Dixon, T., & Jordan, E. (2022). Second language (L2) gains through digital game-based language learning (DGBLL): A meta-analysis. Language Learning & Technology, 26(1), 1-25.

Dunkel, P. (1991). Computer Assisted Language Learning and Testing Newbury House.

Gee, James Paul (2005) "Pleasure, Learning, Video, Games, and Life: the projective stance" E-Learning Vol 2, number 3.

Harris, J., Forde, P., Archer, P., Nic Fhearaile, S., & O'Gorman, M. (2006). Irish in primary schools: Long-term national trends in achievement. Dublin: Department of Education and Science.
 Harvard

Hickey, T. & Stenson, N. 2011. Irish orthography: what do teachers and learners need to know about it, and why? Language, Culture and Curriculum, 24, 23-46.

Kincaid, J., Fishburne, R., Rogers, R. & Choissom, B. 1975. Derivation of new readability formulas (automated readability index, fog count, and flesch reading ease formula) for Navy enlisted personnel. Research Branch Report 8–75.

Law, E. L. C., & Sun, X. (2012). Evaluating user experience of adaptive digital educational games with Activity Theory. International journal of human-computer studies, 70(7), 478-497.

Millour, A., & Fort, K. (2020, May). Text Corpora and the Challenge of Newly Written Languages. In 1st Joint SLTU and CCURL Workshop (SLTU-CCURL 2020).

Millour, A., Araneta, M. G., Konjik, I. L., Raffone, A., Pilatte, Y. A., & Fort, K. (2019, May). Katana and Grand Guru: a Game of the Lost Words. In 9th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics (LTC'19).

Napier, K., & Whiskeyjack, L. (2021). wahkotowin: Reconnecting to the Spirit of nêhiyawêwin (Cree Language). *Engaged Scholar Journal: Community-Engaged Research, Teaching and Learning*, 7(1), 1-24.

O'Brien, C. (2022, March 23). Inspectors concerned over quality of teaching, learning in Irish. The Irish Times. Retrieved April 14, 2022, from https://www.irishtimes.com/news/education/inspectors-concerned-over-quality-of-teaching-learning-in-irish-1.4834184

Oxford, R. L. (2006). Task-based language teaching and learning: An overview. Asian EFL Journal, 8(3).

Prabhu, N. S. (1987). Second language pedagogy (Vol. 20). Oxford: Oxford University Press.

Protheroe, N. 2004. Research report: Motivating reluctant learners. Principal 84 (1): 46–48.

Purgina, M., Mozgovoy, M., & Ward, M. (2017). MALL with WordBricks–building correct sentences brick by brick. CALL in a climate of change: adapting to turbulent global conditions–short papers from EUROCALL, 254-259.

Reinhardt, J. (2019). Gameful second and foreign language teaching and learning: Theory, research, and practice. Palgrave Macmillan.

Sanacore, J. (2007). Needed: Critics of literacy education with a more inclusive perspective. *International journal of progressive education*, *3*(1), 29-43.

Schmidt, R. (1992). Awareness and second language acquisition. Annual review of applied linguistics, 13, 206-226.

Sørensen, B. H., & Meyer, B. (2007). Serious Games in language learning and teaching-a theoretical perspective. In DiGRA Conference (pp. 559-566).

Tsai, Y. L., & Tsai, C. C. (2018). Digital game-based second-language vocabulary learning and conditions of research designs: A meta-analysis study. Computers & Education, 125, 345-357.

Uí Dhonnchadha, E. & Van Genabith, J. A Part-of-speech tagger for Irish using Finite-State Morphology and Constraint Grammar Disambiguation. LREC 2006, May 2006 2006 Genoa

Underwood, J. H. (1987) "Artificial Intelligence and CALL" in Modern Media in Foreign Language Education: Theory and Implementation. National Textbook Company.

Von Ahn, L. (2006). Games with a purpose. Computer, 39(6), 92-94.

Vygotsky, L. S. (1978). Mind in Society: the Development of Higher Psychological Processes. Cambridge, MA:

Ward, M. (2015). CALL and Less Commonly Taught Languages: Challenges and Opportunities. Research-publishing. net.

Ward, M. (2017). ICALL's relevance to CALL. CALL in a climate of change: adapting to turbulent global conditions–short papers from EUROCALL, 328.

Ward, M. (2019). Joining the blocks together–an NLP pipeline for CALL development. CALL and complexity, 397.

Ward, M., Mozgovoy, M., & Purgina, M. (2019). Can WordBricks make learning Irish more engaging for students?. International Journal of Game-Based Learning (IJGBL), 9(2), 20-39.

Xu, L., & Chamberlain, J. (2020). Cipher: a prototype game-with-a-purpose for detecting errors in text. In Workshop on Games and Natural Language Processing (pp. 17-25).

Xu, L., Uí Dhonnchadha, E., and Ward, M., (2022). User Experience Study of "Cipher: Faoi Gheasa", A Digital Educational Game for Language Learning and Student Engagement. In ACM GameSys'22, Athlone, Ireland.