CARAMILLA – Speech Mediated Language Learning Modules for Refugee and High School Learners of English and Irish

Emer Gilmartin[1], Jaebok Kim[2], Alpha Diallo[2], Yong Zhao[3], Neasa Ní Chiaráin[1], Ketong Su[1], Yuyun Huang[1], Benjamin R. Cowan[4], Nick Campbell[1]

[1]Trinity College Dublin, Ireland
[2]University of Twente, Holland
[3]Vrije Universiteit Brussel, Belgium
[4]University College Dublin, Ireland

`gilmare@tcd.ie`

Abstract

In the development of Computer-Assisted Language Learning (CALL) modules there is a growing emphasis on the use of evolving technological developments which include text-to-speech technology (TTS) and automatic speech recognition (ASR). These technologies allow for greater spoken interaction between the learner and the computer, which is particularly useful when native speaker input is not readily available and to elicit learner speech in a quasi-natural environment. They also allow for greater learner autonomy, and facilitate the creation of modules based on effective learning activities at all levels to address needs for accuracy as well as fluency. The present paper reports on two modules which integrate speech technology, (1) a spoken dictogloss, or text reconstruction, module and (2) a pronunciation module. These modules have been implemented in JAVA as part of CARAMILLA, a spoken dialogue-based language learning tool, which is based on an earlier prototype, MILLA. The dictogloss module is being evaluated with two distinct learner groups – adult refugees learning English while living in the host country and learners of Irish, a minority endangered language. The pronunciation module is also being evaluated with the adult refugee group.

Index Terms: CALL, spoken dialogue systems, autonomous learning

1. Introduction

In this paper we describe the design and implementation of two Computer-Assisted Language Learning (CALL) modules - a spoken dictogloss and a pronunciation module intended to add to our existing system [1], and as incremental progress towards our goal of building a spoken language dialogue system to provide flexible practice of second language skills via the Internet. We first briefly review relevant background in the CALL literature, and discuss the challenges and opportunities of provision of language learning resources using CALL. We then describe the user groups we are focussing on, outlining their specific needs and discussing how an online system and the modules described in this paper could benefit them. We overview our work to date with the MILLA system, and then describe the design and development of the dictogloss and pronunciation modules carried out at the ENTERFACE workshop in summer 2016. We describe our plans for evaluation and further development of the modules and discuss plans for future work on the system.

2. CALL applications for autonomous language learning

In this work CALL refers to the use of an artificial environment containing tasks and activities which help learners attain their goals of improving language skills. Speech technology offers interesting opportunities for language education. Such technology has now been extended to the use of speech recognition (ASR) and synthesis (TTS) in addressing specific tasks and in the implementations of complete tutoring systems [2]. Technology has long played an important role in language learning with video and audio courses available since the early days of audiovisual technology. Globalisation and migration, coupled with the explosion in personal device ownership, have increased the need and the opportunities for well designed, pedagogically sound CALL applications.

Many current CALL applications address the receptive skills (listening and reading) with activities much like a traditional listening or reading exercise transferred to a screen [3]. Such exercises offer few advantages over tradition pencil and paper activities with the possible exception of those which include an audio dimension. Pronunciation tutoring applications range from 'listen and repeat' exercises without feedback or with auto-feedback (the learner hears a recording of their attempt) to more sophisticated systems where the learner's utterance is compared with the target utterance and feedback is given on errors and strategies to correct those errors. There are interesting examples of speech-technology-based spoken production training where phoneme recognition is used to provide corrective feedback on learner input. These include CMU's Fluency [4], and Cabral et al's MySpeech [5]. Much effort has been put into creating speech activities which allow learners to engage in spoken interaction with a digital conversational partner. This engagement is considered the most difficult competence for a learner to acquire independently. An effective approach to providing practice in spoken conversation (or texted chat) is to use relatively simple chatbot systems based on pattern matching (e.g. Pandorabots) [6, 7]. Dialog systems using written text and more recently speech have long been used to tutor learners in science and mathematics [8, 9]. However, conversation poses challenges as success does not solely depend on a lexical 'right answer' but rather on the degree to which the learner manages competent spoken interaction. Early language learning systems such as VILTS [10] presented tasks and activities based on different themes which were chosen by the user, while other systems concentrated on pronuncia-

tion training. Recently interest has grown in more holistic systems where learners engage in spoken interactions relevant to their learning needs such as service encounters or job interviews, particularly for assessment purposes [11]. While many of these systems are being developed with an emphasis on the assessment of communicative skills, and greatly aid fluency and pragmatic competence, there is still scope for activities which combine speech technology with learning tasks which help foster learners' acquisition of lexical, syntactic and phonetic accuracy. With improvements in technology it is now possible to provide a range of activities from free conversation practice to more focussed activities. The use of gamification in educational software is currently receiving a lot of attention as a method of increasing learner motivation [12]. This approach lends itself very well to language learning activities, such as 'grammar games' and 'information gap exercises', and has long been a successful element of the language classroom. In CARAMILLA we aim to combine engaging activities which address specific skills and free conversation practice to allow learners to integrate their linguistic skills.

3. Motivation for System

Many language learners today do not follow traditional learning paths, but rather learn when and where they can. For these learners the goal is often communicative competence in the target language for practical purposes. Fundamental challenges are the development of spoken interaction skills, and the integration of lexical and syntactic knowledge to produce accurate and appropriate language in different contexts and to comprehend vocabulary and syntax. In more formal language provision, conversation classes and practice with peers or native speakers are the traditional methods for development of spoken language proficiency, while syntax, lexicon, and pronunciation development needs are met by teacher led instruction. However classes are expensive and learners may not have access to native speakers willing to practice with them. Indeed, even in traditional classroom settings, pronunciation is the poor relation, and although acqui accurate pronunciation and intonation can be greatly accelerated by focussed practice, this seldom forms an integral part of the curriculum.

For practical reasons, language learning providers group learners according to an average of measures of their competence on tests of the traditional four skills - reading, writing, speaking, and listening. The fifth skill, spoken interaction, is often tested in a brief interview. This method of grouping learners by level does not address the fact that many learners, and particularly those who have not worked through a traditional Western modern languages curriculum in a formal learning environment, have 'spiky' profiles, with some skills present at much higher or lower levels than others. Access to independent learning resources which allow study and practice of the particular skills at levels suitable to the individual offers several advantages – learners can improve their weaker skills while learning at times and places convenient to them. An important advantage is that such resources can be offered online free or at very low cost to users – many of the learners most in need of language skills for work, education, and training, or integration cannot afford private bricks and mortar language tuition and state provision is often rudimentary at best.

The text (both spoken and written) underpinning language learning resources has usually focussed on topics common to the widest variety of potential learners possible - pop culture, uncontroversial general or human interest stories, sport, and activities useful to those learning a foreign language for tourism purposes. Specialised texts do exist for groups learning a language for use in fields such as business or medicine, but these are narrowly focussed on the target area. Applications where the underlying text can be pulled from a wide variety of sources suitable to a learner's particular needs can be very helpful to many different learner groups. For an adult migrant, lessons based on texts drawn from social services websites provide the same basis for linguistic or functional learning goals as stories about pop stars, but also motivate learners as there is a clear link to learners' everyday communicative needs. Similarly, for high schoolers, texts based on school subjects in the target language provide relevant vocabulary and content and language integrated learning.

For this project, we focus on two learner groups - adult refugee ESOL (English for Speakers of Other Languages) learners and learners of Irish.

3.1. Language needs of adult refugees

For an adult newcomer, living in a new country presents many challenges. Social, legal and cultural norms may differ greatly from those of the country of origin. When the language of the host community is unfamiliar, these challenges intensify [13]. Without knowledge of the language, access to training and employment is hampered, – employment and earning levels for US immigrants have been shown to be strongly linked to English language ability [14]. Enabling immigrants to acquire basic knowledge not only of the language but also of the customs and daily life of the host country is recognised as essential to successful integration. (Council of Europe, Common Basic Principles of Integration, 2001).

Adult refugees are a particularly diverse group of language learners. As an illustrative example, in Ireland's national language and integration training organisation for refugees in 2008, 93 nationalities were present and educational backgrounds varied widely; while some learners held third level qualifications, others had had little or no formal education. Levels of literacy varied with learners presenting with no literacy in their own language, with literacy in a non-Latin alphabet, and fully literate in the Latin alphabet. The length of time that learners had lived in the host county (Ireland) ranged from several years to a few weeks, resulting in different levels of familiarity with Irish life and the English language. In terms of language ability, profiles were very spiky, with some learners having attained enough English to get by, often appearing fluent but with very low accuracy although listening comprehension could often be excellent. Others had studied formally but had comparatively low spoken interaction skills although reading and writing was of a high standard [15].

Learning a language to live in a country where the language is spoken is not a simple matter of attaining an academic understanding of the language. Better results can be expected when the texts used are tailored to practical communicative needs or areas of interest - a parent would benefit far more from an exercise on Present Tense structures based on the local education system than on a description of life on the Space Station. In addition, free or low cost language learning resources which can be accessed from home at any time are particularly helpful to this group. Activities which

allow topics of interest to the learner to be selected as underpinning text can also foster learner autonomy and motivation [16]. Caramilla's pronunciation and dictogloss activities are designed to be flexible, with pronunciation practice available for many first languages, while the dictogloss can be adapted to any text of interest or level of competence, and thus should be useful to this group. The system is currently being trialled with learners around the CEFRL B1 level, defined as a level where the language learner 'can understand the main points of clear standard input on familiar matters regularly encountered in work, school,leisure, etc. (Council of Europe, 2001). This level was chosen as it is commonly defined as the 'threshold' for integration and relevant to all newcomers. Of course the modules can also be tailored to learners who need to attain higher competences to meet gatekeeping requirements for entry to higher education or to progress towards citizenship.

3.2. Irish language learners

Irish is a Celtic language which was spoken widely throughout Ireland up to the 19th century but has since receded as a community language into just a few small geographical pockets known as Gaeltacht areas. In these areas Irish has extended as a native language in an unbroken chain. Most speakers of Irish, however, have acquired the language through schooling. It has the status of being one of the two official languages in Ireland and since 2007 has recognition as an official language of the European Union. It a compulsory subject of study in all primary and post-primary schools in the Republic (for students from 4-18 years old) but motivation and attitude amongst the learners is very variable [17, 18]. Irish, nonetheless, is classified as an endangered language by UNESCO [19]. There are few native speaker models available to the vast majority of learners of Irish and, anecdotally, the language competence of teachers seems variable as most are themselves second language learners.

Work on speech technology development for Irish has been underway in the Phonetics and Speech Lab., Trinity College, Dublin for some years [20]. Synthetic voices representing the three main dialects of Irish have been developed as part of the ABAIR initiative and are freely available at www.abair.ie. A significant strand of the ABAIR initiative has been the development of CALL materials for Irish [21, 22, 7] .

As part of the CARAMILLA project, the spoken dictogloss game has been designed for Irish. It is intended to be a motivational and challenging CALL tool which goes beyond the range of materials currently available for the teaching of Irish insofar as it gives learners the freedom to choose their own learning materials while integrating speech technology. It is a highly interactive, task-based language learning exercise, which gives freedom to learners to choose topics or themes which are of relevance to them while providing them with access to native speaker models. It is intended to have positive effects on the attitude and motivation of Irish learners at second level as well as being a pedagogical tool for the continuing linguistic development of Irish teachers.

4. CARAMILLA System

The CARAMILLA project at the Enterface 16 workshop, held at the University of Twente's Design Lab, was a follow on to an earlier project (MILLA) at Enterface 14, held in Bilbao. MILLA (Multi-modal Interactive Language Learning Agent) is a dialogue system providing spoken social chat at different levels (two speech-enabled web-based Pandora chatbots), pronunciation and traditional grammar training [1]. MILLA was created in Bilbao by a team of nine.

The goals of the 2016 project at Twente were to produce language learning modules to extend the capabilities of the MILLA/CARAMILLA language learning agent system. There was an onsite team of four postgrad students, who designed and implemented a language game module (dictogloss), and redesigned and implemented an improved pronunciation module.The CARAMILLA project focused on creating a speech enabled version of an engaging and adaptable activity, dictogloss, which integrates all skills in a focussed game, and on redesigning and improving the system's pronunciation training module. These activities and the original MILLA are currently being integrated with a more robust dialog platform (CARA) to form CARAMILLA. The system targets two user groups with differing needs – school language learners and teachers of Irish, and adult refugee learners of English living in Ireland.

4.1. Dictogloss

A dictogloss is a well-known text reconstruction game, which is ideal for implementation in a CALL system as it can be played by one learner with a tutor (the system) or between a group of learners with a tutor. This section describes the design and implementation of the spoken dictogloss module for the CARAMILLA system.

4.1.1. Dictogloss Game Description

A dictogloss is a complete cloze text reconstruction exercise. It is widely recognised as a useful language learning game and the version implemented here is close to that described in Rivolucri's seminal 'Grammar Games' [23]. In the basic form of the activity, learners are exposed to a text, either by reading or listening, and then reconstruct some or all of it - by filling in blanks on a worksheet or by writing and rewriting fragments to assemble a coherent text. The exercise aids acquisition of syntax and vocabulary, and when performed orally, also aids in developing listening skills. A major advantage of the exercise is its flexibility - different areas of vocabulary and syntax can be addressed at different levels of difficulty by simply changing the text used in the exercise. Thus, the game can be very useful in context and language integrated language learning scenarios, where learners learn the language by learning about something else - for example, for refugees learning the past tenses while reconstructing an account of the life of an important figure in the history of the host country drawn from a history resource or a current news story of interest scraped from a news website.

The typical procedure for a spoken dictogloss exercise is as follows:

1. Teacher asks learners to relax, clear desks, and listen - there is no note-taking allowed

2. Learners listen as text is read at normal speed with normal intonation

3. When text is finished, learners write down words remembered from text on a scrap of paper

4. Teacher distributes a blank rectangular grid containing enough cells for all the words (tokens) in the text, or draws the grid on the board.

5. Learners take turns guessing words that were in text, teacher tells them grid co-ordinates and learners enter correct guesses into grid. Points are awarded for correct words.

6. Teacher provides support in the form of hints and encouragement as learners complete grid

The game has two distinct phases - in the first few turns the learners use the words they have remembered and written down. After they have exhausted the words from their wordlists, learners start to reconstruct the sentences in the text by inferring what possible words could fill the blanks in the emerging text. Seeing the string 'I __ a __ car', they will (implicitly or explicitly) realise that a verb is needed in the first blank and will start guessing verbs appropriate to the world of cars. The second blank will elicit adjectives. This hypothesis and test stage is the core of the exercise allowing learners to consolidate their syntactic and lexical knowledge in an interactive context. The constant rehearsing and hypothesising of possible phrases based on the known words activates latent vocabulary and with immediate feedback on guesses the learner refines their syntactic knowledge in the context of the spoken language rather than in isolated grammar exercises.

Added motivation can be provided by grouping learners into teams, playing against a clock, or the use of more detailed scoring systems to reward good guesses - when a learner proposes a word which is the correct part of speech for example. It is clear that several of these additions can be easily incorporated into an automatic version of dictogloss, with virtual players or additional characters, access to NLP tools such as POS taggers, and the use of scoreboards. The spoken dictogloss was implemented in JAVA as part of CARAMILLA as described below.

4.1.2. JAVA Implementation of Dictogloss

The dictogloss process was implemented in Java, by creating web pages using Java Servlets and JSP. Both an English and Irish version were implemented differing only in (1) where the text was retrieved from, in this case the Simple English and Irish Wikipedia pages, and (2) in the text-to-speech (TTS) systems used to read the text to the user. Simple Wikipedia was chosen as the language used is more suitable for learners in terms of lexical and syntactic complexity. Importantly, the sentences are short enough to sound like plausible spoken text, unlike standard Wikipedia articles where collaborative editing often leads to very convoluted embedding and lengthy sentences. The TTS system used for the English version was CereProc's Caitlin (Irish-accent English language) [24], which provides a speech model relevant to learners living in Ireland. For the Irish version, the ABAIR synthesiser was used. ABAIR offers a number of voices representing the three main dialects of Irish, and is an ongoing initiative of the Phonetics and Speech Laboratory, Trinity College Dublin [25]. Thus, CARAMILLA can provide the learner with access to models of the dialects of Irish, which differ considerably. In CARAMILLA, for both English and Irish, we send TTS queries to web servers in Trinity College Dublin to get the synthesis results. Since we have very long texts for the dictogloss game module, lags can be introduced by sending the complete TTS query to the web server. In addition, if we send the whole text content for synthesis, there will be no significant pause between sentences, which sounds unnatural and makes comprehension difficult. Therefore, we split the long text into single sentences, which are sent for synthesis sequentially, resulting in less lag. This method has the serendipitous advantage of creating natural sounding pauses in the synthesised reading of the texts.

In this implementation of dictogloss, the game is browser-based and is accessed from CARAMILLA's main menu. On the menu page, the user selects dictogloss and chooses whether they want to listen to or read the text and then chooses a topic from a currently predefined list. In the future the list could just be a list of subjects that may be of interest to the user and when they choose the subject current articles could be scraped from news websites or journals and returned. Once the user has chosen an action (read or listen) and a predefined topic, the text is scraped from the English or Irish Wikipedia page and a predefined number of sentences are formed for the user. If the user chooses the read option, text will be displayed for them, but if they chose to listen, the TTS system will be called. Once the user has finished reading or listening, they are taken to a page where a grid for the text is presented, as a series of underscores representing words separated by spaces, with the punctuations from the original text already filled in, indicating where sentences end. On this page the user has a textbox in which they can enter words that they think is in the text, modelling the scribble page used in the classroom scenario. If the word that the user types, regardless of case, is in the text, the word replaces the underscores in all the places it appears in the original text. When the user guesses the correct word they get points for the word. Currently each word is only worth a single point, but in the future words could have different points values based on difficulty. While the user is going through and inferring which words come next, they may get stuck and if this happens the user has the chance to ask for help just as they would if they were going through the process with a teacher. By entering in the text 'hint' the user is currently given the missing word, but this is being extended to allow the system to give the user the definition of the word or a synonym of the word. In real life, added motivation can be provided by grouping learners into teams, playing against a clock, or the use of more detailed scoring systems to reward good guesses - when a learner proposes a word which is the correct part of speech for example. It is clear that several of these additions can be easily incorporated into later versions of dictogloss, such as including virtual players or scoreboards.

4.2. Implementation of Pronunciation Module

The pronunciation training module is based on the use of automatic speech recognition to compare a learner's production of an utterance with a model. The intention is to provide pronunciation training for learners of different first languages (L1), as different L1 speakers are known to make characteristic errors.

4.2.1. Corpus of typical errors for pronunciation module

In order to feed our GOP module with some data, we created a pilot corpus of 50 practice sentences with common pronunciation mistakes in English by language background. Pairs of frequently confused English language phonemes for learners from 10 first languages (Arabic, Chinese, Croatian, Dutch, Finnish, French, German, Korean, Spanish, and Turkish) were collected from an expert knowledge website [26]. A carrier sentence for each phoneme pair was then created for use in the application. For each sentence, we created tips for how to

pronounce the commonly mistaken phoneme pair, including a visualisation of the vocal tract, and a phoneme transcription with lexical stress from the CMU pronouncing dictionary [27]. For example, for Spanish or French speakers, the phrase 'These shoes fit my feet' containing the **I vs i:** sounds in the minimal pair 'fit' and 'feet' was used to test these commonly confused vowel sounds.

4.2.2. GOP

Pronunciation scoring is generally utilised in language learning applications to obtain global scores which tells overall goodness of proficiency on an utterance. However, global scores do not give specific information of where students make mistakes, which results in less useful applications [28]. Hence, pronunciation testing should give not only global scores but also local scores at phoneme levels, and learner should pay attention to which phonemes they cannot pronounce correctly.

One method to detect phoneme level errors is the Confidence-Measure (CM) based error detection that often uses Hidden Markov Model (HMM) [29]. A practical advantage of this approach is to utilise an Automatic Speech Recognition (ASR) system using speaker adaptation techniques such as Maximum Likelihood Linear Regression (MLLR). However, the CM-based approach uses the same feature set for all the phones, which might not be an optimal approach to detect explicit errors on a specific phoneme. Moreover, Speaker adaptation can bring over-fitting to a target user, which makes confidence scores even more unreliable. In contrast, Linear Discriminant Analysis (LDA)-based classifiers optimise acoustic feature sets for each phoneme, which results in significant improvement of error detection on particular phones [28]. However, their evaluations are quite limited to small Dutch phoneme sets, which cannot be generalised.

Our approach is based on Witt's goodness of pronunciation (GoP) [29]. Figure 1 describes the basic algorithm. This algorithm calculates a distance between an answer (log likelihood of forced-alignment results) and a target (that of a phoneme recogniser) as follows:

$$GOP_1(q_i) = \frac{\log(p(O|q_i))}{NF(O)} - \frac{\max_{j=1}^{J} \log(p(O|q_j))}{NF(O)} | \quad (1)$$

where q_i is i'th phoneme in an utterance and $NF(O)$ is the sum of log-likelihood of all frames in the observation.
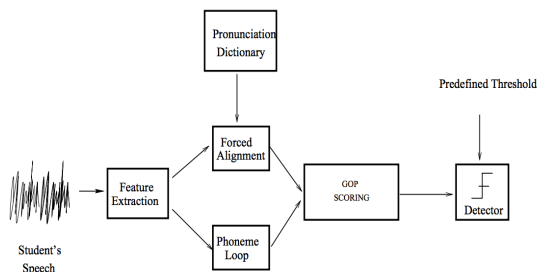

Figure 1: *Basic algorithm of GoP [29]*

Again, this approach can measure a local score for each phoneme but it does not indicate explicit errors (e.g. b → v).

Hence, we modelled explicit error networks for mother languages (e.g. Chinese, Korean, Arabic, and etc.). For this, we collected utterances where non-native speakers make common mistakes from a user group during the eNTERFACE workshop. Figure 2 depicts an example of the error networks.
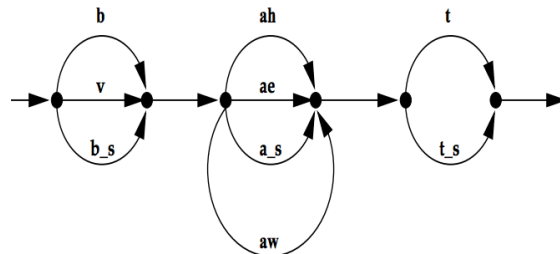

Figure 2: *An example of error model [29]*

The following equation is to combine a local score and a score of its explicit error model:

$$GOP_2(q_i) = GOP_1(q_i) + KGOP_e(q_i) \quad (2)$$

where K is a scaling factor. The details to calculate $GOP_e(Q_i)$ can be found in [29]. In addition, our system explicitly displays which type of errors a specific user makes. Errors are categorised into deleted (missing), substitute, inserted as the same way used in the evaluation of ASR system. Hence, the user could realise both their proficiency and explicit errors. For reproduction, we implemented our system using HTK [30] and Sphinx 4[31] tool kits. For the robust recognition in the wild, we employed sub-band OSF-based voice activity detection [32].

5. Conclusion and Future Work

The dictogloss module is currently being extended to incorporate texts of different levels and on different subjects in both English and Irish, and is being tested with real users. The English language version is being piloted with adult refugee learners in a centre for language and integration courses in Dublin, while the Irish language version is being piloted with Irish language students in Trinity College Dublin. The pronunciation system is currently available in English only and is being tested with the refugee learners.

For future work on the two modules, we plan to integrate an animated avatar into the system. For pronunciation tutoring, apart from getting a score from GOP, it will be very beneficial for the learner to learn how to pronounce by seeing the lip movement from a virtual agent, provided that a suitably accurate avatar can be used. Moreover, when students are playing the dictogloss game, having a virtual agent acting as tutor to read the story and even to talk with them may provide motivation and add to engagement - this is a question we plan to explore, as is that of adding a second agent as a competing or collaborating 'classmate'. We are exploring the open source animation toolkit Smartbody [5] to build a virtual avatar for our CARAMILLA system.

We are currently porting the existing functionality of the MILLA system to CARAMILLA. MILLA already contains two chatbots – a male character at beginner level and a more advanced female character providing free conversation practice.

MILLA's existing user record system, conversation, grammar and pronunciation modules will, combined with the two new modules, will result in a comprehensive learning environment. On the curriculum side, the plan is to eventually provide enough content to allow learners to complete an online portfolio of communicative tasks, modelled on the European Language Portfolios [33]. This portfolio, with learning objectives matching the requirements of the CEFRL, will provide a dynamic record of the learner's competence, both motivating reflection and self-assessment leading to further learning and linking progress to an internationally recognised assessment paradigm.

We hope that the system can eventually be used as a free web-based resource for language learners wishing to learn the English language needed to live and work in Ireland, and by learners of Irish worldwide.

6. Acknowledgments

7. References

[1] J. P. Cabral, N. Campbell, S. Ganesh, E. Gilmartin, F. Haider, E. Kenny, M. Kheirkhah, A. Murphy, N. Ní Chiaráin, T. Pellegrini, and others, "MILLA Multimodal Interactive Language Learning Agent," in *DialWatt - Semdial 2014*, 2014.

[2] M. Eskenazi, "An overview of spoken language technology for education," *Speech Communication*, vol. 51, no. 10, pp. 832–844, 2009.

[3] K. Beatty, *Teaching & researching: Computer-assisted language learning.* Routledge, 2013.

[4] M. Eskenazi and S. Hansma, "The fluency pronunciation trainer," in *Proceedings of the STiLL Workshop*, 1998.

[5] J. P. Cabral, M. Kane, Z. Ahmed, M. Abou-Zleikha, E. Székely, A. Zahra, K. U. Ogbureke, P. Cahill, J. Carson-Berndsen, and S. Schlögl, "Rapidly testing the interaction model of a pronunciation training system via wizard-of-oz." in *International Conference on Language Resources and Evaluation (LREC)*, 2012, pp. 4136–4142.

[6] R. S. Wallace, *Be Your Own Botmaster: The Step By Step Guide to Creating, Hosting and Selling Your Own AI Chat Bot On Pandorabots.* ALICE AI foundations, Incorporated, 2003.

[7] N. N. Chiaráin and A. N. Chasaide, "Chatbot technology with synthetic voices in the acquisition of an endangered language: Motivation, development and evaluation of a platform for irish." in *LREC*, 2016.

[8] A. C. Graesser, P. Chipman, B. C. Haynes, and A. Olney, "AutoTutor: An intelligent tutoring system with mixed-initiative dialogue," *Education, IEEE Transactions on*, vol. 48, no. 4, pp. 612–618, 2005.

[9] D. J. Litman and S. Silliman, "ITSPOKE: An intelligent tutoring spoken dialogue system," in *Demonstration Papers at HLT-NAACL 2004*, 2004, pp. 5–8. [Online]. Available: http://dl.acm.org/citation.cfm?id=1614027

[10] M. E. Rypa and P. Price, "VILTS: A tale of two technologies," *Calico Journal*, vol. 16, no. 3, pp. 385–404, 1999.

[11] K. Evanini, S. Singh, A. Loukina, X. Wang, and C. M. Lee, "Content-based automated assessment of non-native spoken language proficiency in a simulated conversation."

[12] J. P. Gee, "Game-like learning: An example of situated learning and implications for opportunity to learn," *Assessment, equity, and opportunity to learn*, pp. 200–221, 2008.

[13] D. G. Little, *Meeting the language needs of refugees in Ireland.* Refugee Language Support Unit, University of Dublin, Trinity College, 2000.

[14] B. R. Chiswick and P. W. Miller, "Immigrant earnings: Language skills, linguistic concentrations and the business cycle," *Journal of Population Economics*, vol. 15, no. 1, pp. 31–57, 2002.

[15] E. Gilmartin, "Language training for adult refugees: The integrate ireland experience." *Adult Learner: The Irish Journal of Adult and Community Education*, vol. 97, p. 110, 2008.

[16] D. Little, "Language learner autonomy: Some fundamental considerations revisited," *International Journal of Innovation in Language Learning and Teaching*, vol. 1, no. 1, pp. 14–29, 2007.

[17] L. Murtagh, "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," 2003.

[18] N. N. Chiaráin, "Text-to-speech synthesis in computer-assisted language learning for irish: development and evaluation," Ph.D. dissertation, Trinity College Dublin, 2014.

[19] C. Moseley, *Atlas of the World's Languages in Danger.* Unesco, 2010.

[20] A. N. Chasaide, N. N. Chiaráin, C. Wendler, H. Berthelsen, A. Murphy, and C. Gobl, "The abair initiative: Bringing spoken irish into the digital space," in *Proceedings of Interspeech 2017*, 2017.

[21] N. N. Chiaráin and A. N. Chasaide, "The digichaint interactive game as a virtual learning environment for irish," *CALL communities and culture–short papers from EUROCALL 2016 Edited by Salomi Papadima-Sophocleous, Linda Bradley, and Sylvie Thouësny*, p. 330, 2016.

[22] ——, "Evaluating synthetic speech in an irish call application: influences of predisposition and of the holistic environment." in *SLaTE*, 2015, pp. 149–154.

[23] M. Rinvolucri, "Grammar games: Cognitive," *Affective, and Drama Activatiobn for EFL Students*, 1984.

[24] "CereVoice Engine Text-to-Speech SDK | CereProc Text-to-Speech," 2014. [Online]. Available: https://www.cereproc.com/en/products/sdk

[25] "abair.ie The Irish Language Synthesiser." [Online]. Available: http://www.abair.tcd.ie/

[26] "Teaching and learning English pronunciation." [Online]. Available: http://www.tedpower.co.uk/phono.html

[27] "The CMU Pronouncing Dictionary." [Online]. Available: http://www.speech.cs.cmu.edu/cgi-bin/cmudict

[28] H. Strik, K. Truong, F. De Wet, and C. Cucchiarini, "Comparing different approaches for automatic pronunciation error detection," *Speech Communication*, vol. 51, no. 10, pp. 845–852, 2009.

[29] S. M. Witt, *Use of speech recognition in computer-assisted language learning.* University of Cambridge, 1999.

[30] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, D. Povey *et al.*, "The htk book," *Cambridge university engineering department*, vol. 3, p. 175, 2002.

[31] W. Walker, P. Lamere, P. Kwok, B. Raj, R. Singh, E. Gouvea, P. Wolf, and J. Woelfel, "Sphinx-4: A flexible open source framework for speech recognition," 2004.

[32] J. Ramírez, J. C. Segura, C. Benítez, A. De la Torre, and A. Rubio, "An effective subband osf-based vad with noise reduction for robust speech recognition," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 6, pp. 1119–1129, 2005.

[33] D. Little, "The common european framework and the european language portfolio: Involving learners and their judgements in the assessment process," *Language Testing*, vol. 22, no. 3, pp. 321–336, 2005.