

Second Language Learning Using Bilingual eBooks

Tiago Soares¹, Xavier Anguera², and Isabel Trancoso³

¹ Instituto Superior Técnico / INESC-ID tvdsouares@gmail.com

² ELSA Corp. xanguera@gmail.com

³ Instituto Superior Técnico / INESC-ID Isabel.Trancoso@inesc-id.pt

Abstract. Current Computer Assisted Language Learning (CALL) tools, which have largely remained unchanged over the past few years, are often inadequate for attaining fluency in a second language. The goal of this work was to build a CALL tool based on bilingual eBooks, following the EPUB3 format, which aims to develop a wide array of necessary competences required from language learning. The tool should have an intuitive user interface, include the possibility of retrieving word definitions, and also include a set of gamified exercises.

The project comprised the development of a Python script which took the raw inputs (parallel texts in L1 and L2 and pre-recorded audio in L2) and merged them into a single structured eBook document, complete with dynamic features. These features offer the user an improved control over the audio playback (for instance, its speed), and are not included by default in the target platform, Apples iPad in this work. The script also retrieves, from an open-source API, dictionary definitions for every word in the eBook and includes them in the final document. Additionally, the project also comprised the development of a Javascript application that implements the gamified exercises included in the eBook, without relying on online external resources. This solution was required, since the chosen eBook format specification does not allow for online content to be fetched from inside the eBook.

A preliminary user study was conducted with a demonstration eBook generated by this system in order to improve its features and understand the expectations users would have for the eBook. Although positive feedback reassures the usefulness of this application, the demand for features that require deeper analysis of the dual language book raises concerns over the viability of the chosen reading platform, Apples ibooks application, and the EPUB3 format versus implementing the solution as an independent, standalone mobile application.

1 Introduction

Effects of globalization in recent years have created a demand for bilingual and multilingual individuals. Several languages such as English and Spanish have become highly sought after by employers for their value in international communication, the tourism industry and in trade. Other emergent languages, such

as Mandarin Chinese, are also sought after for the value it is perceived of them from times to come. Migration has also influenced these demands as it shapes countries into multilingual and multicultural societies, and so has the Internet with its accessibility to readily connect individuals anywhere around the globe. It is for these reasons that people and countries as a whole seek to prepare themselves to deal with the challenges of a modern bilingual world by instructing themselves and their youth in a second language (L2).

Despite the vast accessibility to foreign media, which supplies a plethora of means of study and opportunities for enriching oneself with new knowledge of the language to advanced students, most of the tools available to beginner and intermediate L2 learners have remained unchanged over the past few years. These instruments are generally textbooks, audio compact discs and other similar media, which allow students to study and learn some new insights into the L2 on their own. Proficiency in L2, however, comes from achieving a degree of fluency, complexity and accuracy. These are attained from acquiring a competence in understanding formulaic expressions and knowledge of the L2 grammatical rules [11]. For a student to obtain these skills, studies suggest that instruction should revolve around an extensive input in the L2, as well as focus on the meaning of its formulaic expressions, but also give the student the possibility of output [5].

This paper describes the development and evaluation of a tool to generate multimodal, bilingual ebooks in EPUB3 format, which in addition include embedded gamified exercises to train L1 speakers onto L2 language.

2 Background

2.1 Theories for language learning

Computer-assisted language learning CALL is an approach to language learning and teaching which includes computer technologies as a means to promote educational learning through a substantial interactive element. It can be used to help students attain fluency, complexity and accuracy in oral skills, literacy and formulaic expressions [6, 8].

Owing to audio features, CALL programs provide their users with a profusion of different ways in which they can practice their listening skills. For instance, videos supporting L1 or L2 subtitles or explanatory notes, along with an option to quickly and easily control its playback can improve both immediate comprehension and acquisition [3].

Also, thanks to ASR (automatic speech recognition) systems, L2 learners are no longer restricted to practice speaking with a partner, but are able to talk to a computer to evaluate their performance, or practice short dialogues [12, 9, 1].

For writing skills, L2 learners can use CALL programs to enrich their vocabulary and benefit from the now widespread spell and grammar checking systems. Such systems enable L2 learners by quickly giving them feedback which develops the L2 writing accuracy of the learner.

Gamification Gamification is the concept of applying game design mechanics, techniques and game principles to improve user engagement, productivity and motivation. Games require the players to learn, even if just the rules of the game. Upon mastering the rules, the player has the creative freedom of applying and navigating the rules. When developing a new game, the game designer must take into consideration how to convince the player to learn how to play the game and use good methods and principles to ensure that it is a fun experience [10].

2.2 Text and audio processing algorithms

Parallel text alignment When dealing with large parallel texts, such as bilingual eBooks, the task of alignment can become challenging. When a text is translated, particularly when considering how different two languages can be in terms of structure, grammar and syntax, some sentences may be split, merged, reordered, deleted or inserted by the translator.

Alignment may be executed at different levels. Some texts may benefit from paragraph alignment, while others from sentence alignment or even word alignment. Regardless, alignment methods are often categorized as either statistic or lexicon-based.

One such statistic approach is that of Gale and Church [7]. The motivation for their model came from the observation that longer regions of text tend to have long translations, and shorter regions have shorter translations, finding that the correlation between the length of a paragraph, measured in number of characters, between the original text and its translation was exceedingly high.

Lexicon-based alignment approaches use associative measures. Most basic lexical level methods use the Dice-coefficient, which is a statistic used for comparing the similarity of two elements, to keep co-occurrence scores for each word pair, and the total independent occurrences for those words in each of their respective texts. Then it calculates the probabilities of alignment and selects the sentence combinations of maximum likelihood.

A combination of both statistic and lexicon-based alignment methods is the approach of most modern parallel text aligners, such as the one used in the present work, Hunalign, an open-source parallel text aligner that aligns text on the sentence level.

Forced alignment and speech recognition systems Forced alignment systems attempt to match the words of a given input text to those of an input audio file. The input text is mapped into a phone sequence by using a pronunciation dictionary, and phone boundaries are determined by algorithms such as the Viterbi algorithm.

Current forced alignment systems will usually be composed of two parts, a training stage and an alignment stage. During training, the system learns the correspondence between phonemes (usually including left-right context) and audio-derived features.

When attempting alignment, the system aligns the new speech with its acoustic models through the Viterbi algorithm to produce time stamps for each word

in the given audio file. The forced alignment process is identical to that of ASR, except that a phonetic transcript is known [4]. Thus, in ASR the Viterbi algorithm is used to find the most likely path through the phoneme HMM and from that the system then generates the resulting word string.

In this work we used a slightly different approach proposed in [2] where both text and audio are converted into phoneme sequences and then a dynamic programming algorithm based on the edit distance aligns both sequences to obtain the alignments. This method has the advantage over standard forced alignment techniques in that it can align very long audios to texts without the usual memory burdens of decoding long audio files.

2.3 EPUB 3.0.1

EPUB is a widely adopted format for digital books, eBooks, due to its potential to comprise complex layouts and interactivity in the eBook. This is achieved by the use of the latest HTML5 standard by EPUB 3.0.1, which allows EPUB publications to contain video, audio and interactivity through scripting, much like modern web pages. EPUB reading systems paginate the content dynamically, adapting it to the display, rather than having the reader zoom in or out a fixed format page, as is common with other applications such as PDF. This behavior is not forced in any way, however, allowing for specific content that would benefit from a zooming and panning functionality to be manipulated in such a way.

Being comparable to a web page EPUB is, consequently, a simple ZIP archive file that contains the HTML files, CSS style sheets, metadata and any other assets necessary for EPUB reading applications or devices. One such file contained in the EPUB is the Package Document, which is a unique file in the publication. It specifies all documents, images and other content that make up the publication, including the ordering of the pages of the eBook.

Most EPUB readers support the use of the synchronized multimedia integration language (SMIL). SMIL allows the synchronization of the transcript of a publication with a corresponding recorded audio file. It is embodied by a timings file that matches marked segments of text to portions of the audio file linked to each page.

3 Specification and Design

3.1 Specification

The objective of the software system developed in this project is twofold. First, it must create the eBook and its user interface, complete with a dictionary mode. The user interface must allow control over both playback and speech rate. Secondly, it must include gamified exercises in the eBook. It is intended that the creation of the EPUB formatted eBooks be fully autonomous, requiring no external resources other than the text given in two distinct languages, its corresponding audio file and parallel text alignment and forced alignment tables (which in turn could be obtained automatically using other automated tools).

Due to the use of the EPUB format, any resource or data required by the eBook must be self-contained, as it is not possible for the eBook to fetch any online resource. Static resources available online required by the eBook must be fetched by the eBook generator script.

Regarding the specifications of the parallel book section, the eBook must provide the user with a sentence highlighter which establishes a correlation between the sentence being read by the narrator and the corresponding text segment in the second language. The user interface must allow the reader to seamlessly control audio playback, through commonplace play and pause functions and pace and flow control processes as well.

The dictionary mode must be implemented separately as the specifications for this project render the reader's own dictionaries inaccessible. This, however, enables future development to disambiguate the word definitions. The dictionary mode is expected to be able to return the definition of every word in the eBook that has an entry in the given source. A few select words are to be underlined at all times, meaning their definitions may be accessed without the use of dictionary mode. These correspond to the words deemed most important and relevant for the readers learning process.

Regarding the gamified exercises, there are a few restrictions imposed on their design. The games revolve around the text in the eBook, requiring the user to translate words into the second language given their context in the original transcript. The games are also expected to utilize gamification elements to bolster the users motivation and engagement levels.

3.2 Design

The eBook is to allow the user to play the audio starting from any given sentence in the eBook. To do this, the user should tap the text corresponding to the sentence he wishes to start from. In addition, the eBook is to provide means that help better comprehend the audio. It is the responsibility of the EPUB generator script to include this functionality in all eBooks. On running, this first component expects a number of eBook independent resources, namely icons, images, text Fonts, scripts and optionally a *list of important words* to underline in the eBook that is going to be processed. The images and icons are needed to make a small graphical user interface, which is required to be non intrusive, so as to not disrupt reading.

In addition to the dictionary mode, the user interface includes tools which ease reading along with the audio. These functions are displayed in figure 1. In order, they are play/pause, rewind, repeat, one-by-one mode (in which only one sentence is played, at the end of which the audio is paused), slow/normal pace audio and the dictionary mode. The *important word list* is optional. If none is present, important words are selected from the text on an occurrence frequency basis. These words are to be underlined by default on the eBook, and pressing them will bring up a short definition. Regardless of importance, however, every word in the eBook will have a corresponding definition which can be accessed through the dictionary mode in the user interface.

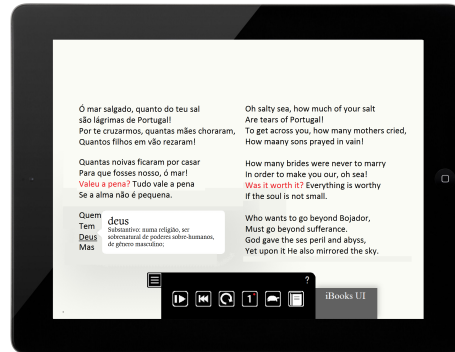


Fig. 1. Graphical user interface and word definition.

The gamified exercises module is placed in the last page of the eBook. Upon loading this page the user is greeted with two games to choose from, *Fill in the blanks* and *Word blitz*. The execution of these games is similar, but their presentation is fundamentally different. In both games, the eBook's text is used to create the games. Sentences or whole pages are selected, and from them, a random word or set of words is replaced with a blank which the user is then prompted to fill with the correct missing word, with the assistance of the corresponding audio.

In the first game, *Fill in the blanks*, whole book pages are used, and several words are removed from it. Initially hidden, one by one, the sentences are then revealed to the user, and the corresponding audio track is played. Should a revealed sentence have one or more of these blanks, execution is halted until the user fills all blanks. When the page is completely revealed and all game actions are exhausted, a score is given to the user depending on how they fare. This design results in prolonged games which tests a subject's consistency.

The second game, *Word blitz*, however, uses a single sentence with a single blank per game, resulting in quicker, more relaxed games with which the player may spend any amount of time they see fit. This will both directly and indirectly aid them in *Fill in the blanks*, as playing *Word blitz* translates not only into acquiring a greater closeness to the eBook, but also into hints that may be used to help achieve higher scores.

The final feature, the hint system, is a gamification tool that rewards users for their time spent in self-assessment. By design, the hint system is only usable during *Fill in the blanks*, but hints may be earned by participating in either game. As players correctly identify the missing words in the games, a progression bar fills. Once entirely full, they earn a hint. In order to use the hint, for the current sentence, a blank has to be selected through the hint menu's drop down list. Once the blank is chosen, two types of hints are available, *Check correctness*, which, without detriment to their scores, reveals whether the text written in the blank is the correct solution; and *Partial solution*, which writes an initial portion

of the word in the blank, useful for when the user has trouble ascertaining the exact word in the audio playback.

4 System Implementation

4.1 EPUB Generator

Upon executing the EPUB generator, it loads all resources required for creating the eBook that necessitate no editing, such as the audio files, scripts, images and icons. Then, it loads the parallel alignment and forced alignment tables and uses them to make a word count. Then, depending on the user, a list of words considered important in the eBook is either pulled from a list, or generated with the word count. Should the latter be selected, two parameters decide which words are added to the important word list: a count threshold, and a word length (in characters) threshold.

Subsequently the EPUB generator script loads the text for both languages and, page by page, merges their content together into a single file, complete with header declarations and user interface elements. This process continues by using the parallel alignment table to create identifiers for sentences or sentence groups. These identifiers will later allow the SMIL emulator to highlight corresponding sentences in both languages as they are read by the audio playback. Finally, the input forced alignment table is read and converted from its word by word representation into groupings delimited by the identifiers. A script element is then added to the page containing a data structure which lists all SMIL timing information for that specific page. Once all pages have been processed, the dictionary manager is loaded, and its definitions incorporated into the eBook. This module consists of a class definition for a dictionary manager and database. The methods defined within the class allow for querying Wiktionary's MediaWiki API for new definitions, and parsing the received response. The parser was implemented specifically for this application, as no other parsers seem to exist for Wiktionary.

Finally, the EPUB generator script concludes the task by creating a final HTML page for the gamified exercises, as well as the Package Document.

4.2 SMIL emulator

This module borrowed its backbone from the now deprecated `rb_smil_emulator` Javascript module created by Alberto Pettarin. `rb_smil_emulator` allowed for mirroring the operation of iBooks's built-in SMIL interpreter. Taking this approach was necessary, as each of the eBook's pages must be bound to a single SMIL file, which in turn must be associated to a single audio file. Given there is no support to control the pace of the audio when using SMIL, it was deemed necessary to use two separate tracks, one for the regular audio track, and a slowed down version of the same track. This demands the creation of a separate SMIL table, which in turn rules out the possibility of using iBook's built-in reader.

The final implementation consists of a Javascript code able to not only simulate the iBooks SMIL interpreter but also provide extra features otherwise not

available. Such features include ways to repeat or rewind the audio playback, set a multiplier to the SMIL data and switch the audio track to reproduce the toggle effect for slowing/speeding up the audio playback, and enabling dictionary mode during which no audio playback is allowed.

4.3 Gamified exercises

The first game, *Fill in the blanks*, uses a version of the SMIL emulator mentioned beforehand. It is heavily modified, includes only the normal pace audio file, and does not register touch events. Instead, the task of moving on to the next sentence of repeating the audio of the active sentence falls upon function buttons. The source of contents for this game is a page in the book, randomly selected by the creation tool. The selection of words to be removed (i.e. the blanks) is made before the game begins, and these words are selected randomly within the boundaries of three parameters. Firstly, the distance between two words must be within limits specified by the designer. The final parameter is the minimum word length. Only words greater in character count than this parameter are used in calculating the distances of the previous two parameters. This process uses a cumulative distribution function to ensure that the randomly selected words meet this criteria. All selected words are then stored in a list for use with the hint system and scoring.

On the second game, *Word blitz*, a page is also selected at random. Upon acquiring the number of segments in the selected page, a text segment is chosen randomly, and a single word in it is selected to be the removed. Both these selections have criteria, as in *Fill in the blanks*. Eligible words are those with a character length greater than a given threshold, and a minimum threshold for the segment length, in words, is also used.

5 Results and Evaluation

Feedback from a test group supported the development of the system. The test group comprised of seven individuals from both the United Kingdom and the United States of America who had either negligible or no prior experience with the Portuguese language. This group indicated which features are of their preference and offered further development suggestions. This option was chosen in lieu of a comparison of a test group using the eBooks developed in this project and a control group using a different form of learning aids. The reason for this being that when using small test groups, it is inconceivable that both groups integrate individuals of comparable learning propensity and identical L2 backgrounds. Consequently, it would be impractical to analyze and compare results based on the two test groups.

It should be noted that all participants were shown the same short demonstration eBook. The L1 for the eBook being English and the L2 target language Portuguese, meaning all audio and games are featured in Portuguese. While the text pages are the same for all participants, due to the random nature of the

gamified exercises, no two participants were given the exact same games to play. Even those who coincidentally chose the same game pages did not experience the same challenges, owing to the fact that the blanks are generated dynamically.

Furthermore, a group of fourteen Portuguese individuals participated informally in the user study. Their feedback was taken into consideration for matters such as usability, user interface and user experience development. Their experiences could not be included in the results since they were given the same demonstration eBook, where the target language matched their mother tongue.

5.1 Survey hypotheses

1. **Audio support promotes second language comprehension in dual language books.** This is the main idea behind this work. Every other function expands on this concept and as such, it is considered the focal point of the survey.
 - (a) **The highlighting aids reading along to the audio playback.** It is expected that the offering of a simple visual queue to the reader facilitates reading along to the audio playback.
 - (b) **The highlighting allows for an easier correspondence of phrases in either language.** The anticipated effect is that of aiding in the analysis of the meaning of the corresponding highlighted sentences. Knowing the corresponding blocks of text is expected to enable a finer interpretation of each individual word.
 - (c) **Control over audio speed and playback facilitates the acquisition of phonetic characteristics of the L2.** This encloses all other implemented functions, such as the UI, slower playback speeds, and playback control. One example is that of the replay button, as it allows maintaining the text in plain view while the audio is being repeated, as often as one would like, without having to press the transcript, thus blocking their own view.
2. **Gamified exercises motivate and gratify second language learning.** Gamification, when well designed, should appeal to its target audience and influence their behavior concerning their tasks, raising levels of engagement, motivation and sense of gratification.
 - (a) **The hints system drives users to dedicate more time to their studies.** This acts as the game's currency. It correlates directly with a tangible benefit the students may use at will. It rewards those who carefully study the book beforehand, and eases any possible feeling of stress derived from a situation where one correct answer is all that is holding back the student from a perfect score.
 - (b) **The scoring system encourages players to commit themselves for better results.** As briefly touched upon on the previous hypotheses 2(a), the scoring system should drive players who appreciate gaming self-elements to raised levels of engagement in an attempt to pass the books games with a perfect score.

5.2 Informal discussion

Regarding current functionalities, participants criticized several features. Amongst those were the lack of a toggle or feedback and responsiveness when using certain functions, word definitions being poor and the original audio pace being too quick. In addition, some concern regarding the effectiveness of the eBook for language learning arose due to the fact that these books are very poor in regards to learning correct grammar.

Ideas offered for future development included a word for word highlighter and single word audio playback, the inclusion of IPA phonetic notation in the dictionary and further functionalities to control audio pace.

5.3 Survey results

Participants were surveyed on their perceptions of both the dual language book as well as the gamification elements included in the exercises. All propositions are answerable on a scale of one to five, where one represents a complete disagreement and five a strong agreement.

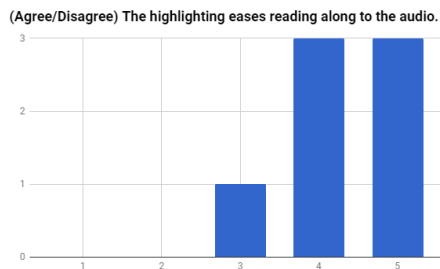


Fig. 2. eBook question 1. Mean: 4.29.

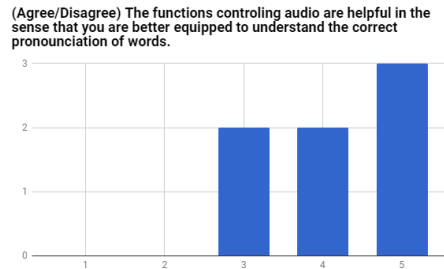


Fig. 3. eBook question 3. Mean: 4.14.

Results regarding this and other questions were largely favorable, indicating a general consensus that the implemented functions are well suited to meet the formulated hypotheses.

Although results for the gamification elements segment of the survey were mostly positive, the hints system pales in comparison to the other features, indicating that this feature requires additional development. Moreover, when filtering the results with respect to demographics, focusing on individuals who did not speak Portuguese, questions 1 and 3 are altered significantly, with predominantly neutral responses.

6 Conclusions and Future Work

The hypotheses formulated in the preceding chapter were met to a satisfactory degree by the results of the user study. The hypothesis 2(a), regarding hints

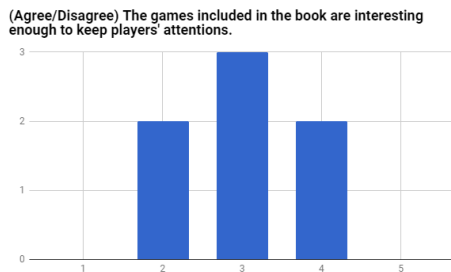


Fig. 4. Gamification elements question 1. Mean: 3.00.

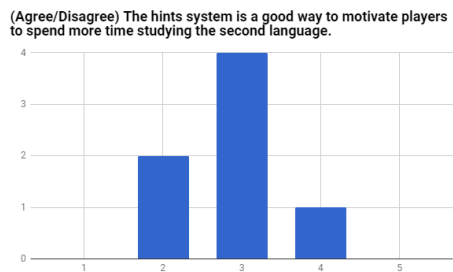


Fig. 5. Gamification elements question 3. Mean: 2.86.

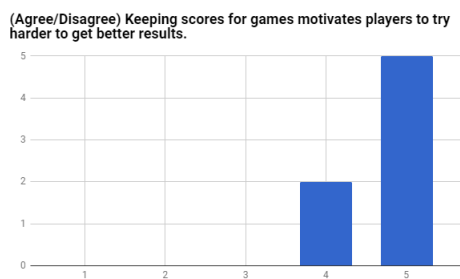


Fig. 6. Gamification elements question 4. Mean: 4.71.

system, however, fell short in delivering the intended gamification element of currency and should be improved upon.

Considering the implementation phase and features participants requested for during the user study, Apple's eBook reader, iBooks, fell short of expectations. It was not possible to use SMIL dynamically as Javascript code that altered the HTML pages in any form ran sluggishly and, despite tailoring a SMIL emulator to suit the eBooks's requirements, timing issues plagued every function call that depended on audio playback.

Considering that the above issues were not at all present when testing on web browsers such as Google Chrome and Internet Explorer, one assumes that iBooks is not prepared for dynamic content not created through its own built-in functionalities. These processes are, however, very basic in nature and unsuitable for projects that require even the least amount of complexity.

Future work should explore options beyond iBooks. As of a year prior to the writing of this article, no other reader applications existed which offered a greater array of developmental tools than iBooks. Perhaps an eBook reader could be designed to catalog and display the eBooks. Having control over the reader application would allow for designs of greater complexity, including features unavailable in other readers, such as loading auxiliary files to meet data requirements for simpler software architectures. Also unavailable in iBooks, internet access could allow for recommendations of eBooks given each students'

personal library or latest readings. Current features that could be improved upon include the dictionary, which would benefit from comprehensive word definitions and their IPA phonetic representation. A grammatical analysis of the text would enable narrowing down the word definitions to those relevant for each context. Moreover, the grammatical analysis would allow for the much-requested grammatical highlighting function during the user study.

Acknowledgement. This work has been supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013.

References

1. X. Anguera and V. Van *English Language Speech Assistant* Interspeech 2016. pp. 1962-1963.
2. X. Anguera and J Luque and C Gracia *Audio-to-text alignment for speech recognition with very limited resources*. Fifteenth Annual Conference of the International Speech Communication Association. 2014.
3. I. Borrs and R. C. Lafayette. Effects of multimedia courseware subtitling on the speaking performance of college students of french. *Modern Language Journal*, 78 (1):61–75, 1994. <http://dx.doi.org/10.2307/329253>.
4. S. Brognaux, S. Roekhaut, T. Drugman, and R. Beaufort. Automatic phone alignment. In H. Isahara and K. Kanzaki, editors, *Advances in Natural Language Processing*, volume 7614 of *Lecture Notes in Computer Science*, pages 300–311. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-33982-0. 10.1007/978-3-642-33983-7_30. URL http://dx.doi.org/10.1007/978-3-642-33983-7_30.
5. R. Ellis. Principles of instructed language learning. *Asian EFL Journal*, 7(3):9–24, September 2005. ISSN: 1738-1460.
6. H. Franco, et al. The SRI EduSpeak™ system: Recognition and pronunciation scoring for language learning. *Proceedings of InSTILL 2000* (2000): 123-128.
7. W. Gale and K. Church. A program for aligning sentences in bilingual corpora. *Computational Linguistics*, 19(1):75–102, 1993.
8. P. Hubbard. *Computer assisted language learning: Vol 1 (critical concepts in linguistics)*. 2009.
9. R. Kanters, C. Cucchiarini, and H. Strik. The goodness of pronunciation algorithm : a detailed performance study. In *In SLATE 2009 - 2009 ISCA Workshop on Speech and Language Technology in Education*, pages 2–5, 2009.
10. D. R. Michael and S. L. Chen. *Serious Games: Games That Educate, Train, and Inform*, chapter 2: Serious Games Defined, pages 17–29. Course Technology PTR, 2006. <http://dl.acm.org/citation.cfm?id=1051239>.
11. P. Skehan. *A Cognitive Approach to Language Learning*. OUP Oxford, 1998. ISBN: 0194372170.
12. S. Witt and S. Young. Phone-level pronunciation scoring and assessment for interactive language learning. *Speech Communication*, 30:95–108, 2000.