Automatic generation of multilingual crossword puzzles with WordNet

Student: Jeanette Tan Yi Wen (U0930152L)

Supervisor: Assoc Prof Francis Bond

HG499 Graduation Project
April 2013

Abstract

This project implements an automatic crossword puzzle generator in Python and a JavaScript solving interface that allows users to tailor the puzzle to suit their preferences. The crosswords are multilingual with the hints and solutions in different languages as it makes use of the Open Multilingual Wordnet which has linked the WordNets of many different languages together. It aims to provide a fun and effective way for language learners to acquire vocabulary.

Keywords: crossword, wordnet, multilingual, automatic generation, computational linguistics, call

Acknowledgements

Francis Bond for supervising this project and for developing the Open Multilingual Wordnet

Takayuki Kuribayashi for providing the readings for kanji in the Japanese WordNet

Saeko Ogihara for her assistance in the recruitment of participants for the study

Bryan Helmig for the Python Crossword Puzzle Generator

Pavel Simakov for the JavaScript Crossword Engine

Contents

List of Tables

List of Figures

1 Introduction

The objective of this study is to implement an automatic multilingual crossword puzzle generator that makes full use of WordNet's capabilities to help language learners master vocabulary effectively. It is targeted at beginning students of the foreign language, as more advanced learners may find monolingual puzzles in their target language more effective.

As a machine-readable dictionary-thesaurus of words interlinked with one another, WordNet has many applications in the field of computational linguistics. This project aims to explore its potential and to facilitate vocabulary acquisition in language learners.

The crossword puzzle is a language game of great educational and entertainment value, and its automatic generation will allow users to customise the puzzle as desired to suit their individual requirements. With regards to its entertainment value, "(the) benefit of using crossword puzzles in the classroom is that they are associated with recreation, and can be less intimidating for students as review tools. Students who might normally balk at practice tests, flashcards, or review sessions with the teacher find puzzle solving to be much less threatening and more like game play." (Alimemaj, 2010)

The educational value of the crossword puzzle lies in its similarity to flashcards as both are essentially paired associate learning tasks, making use of associative memory to recall the corresponding response when a stimulus is presented. The items used in the pairs may not necessarily be single words, they could also be definitions, lists of synonyms, etc. "Studies have shown that in a paired-associate learning task, large numbers of words can be memorised in a very short time (Fitzpatrick, Al-Qarni, & Meara, 2008; Nation, 1980; Thorndike, 1908). Vocabulary learnt in a paired-associate format is also resistant to decay (Fitzpatrick et al., 2008; Thorndike, 1908) and can be retained over several years (Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Bahrick & Phelps, 1987). Recent studies have also suggested that flashcard learning may transfer to normal language use and is a valuable learning activity (Elgort, 2007; Webb, 2002, 2009)." (Nakata, 2011) The

advantage crossword puzzles have over flashcards is that the former is a more active type of rehearsal, requiring users to actually write or type out their answer rather than just reading it, "(engaging) students with the material more than passive types of review techniques do." (Alimemaj, 2010)

Previous research has been conducted on automatic crossword puzzle generation with WordNet (Aherne & Vogel, 2006). This project will take it a step further by introducing multilingual puzzles, made possible by the linking of different languages in the Open Multilingual Wordnet.

The program is written in Python, using CGI and JavaScript to allow users to play in a web browser for greater cross-platform accessibility. It works best in recent versions of browsers like Mozilla Firefox, Opera and Google Chrome.

To clarify the terminology pertaining to crossword puzzles used in this paper:

Solutions are the answers to each item of the crossword puzzle with which players fill in the puzzle boxes

Clues are shown at the side or below the puzzle grid to help players guess the solutions, and in this program they relate to the meaning of the solutions

Hints are not a standard part of the crossword puzzle, but they may be used by players when they give up on guessing the solution through clues alone, and in this program they relate to the word structure of the solutions

Although there are other crossword generation tools on the Internet, they only do the work of placing the words on a crossword grid, requiring users to devise the puzzle themselves by providing a word list. While still allowing users the option of using custom word lists, this project aims to remove that step by making use of WordNet, and also allows users to customise the puzzle in the following ways.

Users may select any two languages from the languages in the Open Multilingual Wordnet as the language pair to be used for the crossword solutions and clues. Although the Open Multilingual Wordnet features a variety of languages and the program is designed to be flexible enough to accommodate all of them, this project will focus on English and Japanese as trying to perfect the user experience for all the other languages would take a lot of time. Solutions in Japanese are written in hiragana instead of kanji, firstly because the short character lengths of many Japanese words when rendered as logograms make them less than ideal in a crossword puzzle, secondly because they are more appropriate for beginners, and thirdly because the Input Method Editors (IMEs) (software solutions to allow foreign languages like Japanese to be typed with a standard alphanumeric keyboard) used on the computer make the selection of the correct kanji a rather simple task.

Users may also choose the type of clues to be displayed, which could be translations, definitions, examples or images, all drawn from WordNet.

This report will describe the creation of the program and its features, and evaluate how it aids in the acquisition of foreign language vocabulary in a small experiment.

2 Program

This section is an overview of the major steps taken to develop different parts of the crossword generation program and solving interface, describing a few of the many changes involved. As suitable scripts were already freely available as open source programs, we decided to adapt the existing scripts to our purposes rather than reinvent the wheel. The modified scripts used in our program will be made available on the crossword website, though they will not be included in this report due to their length. As of writing, the program is accessible to users connected to the NTU network at http://172.21.171.234/~jean0011/cwpz/.

2.1 WordNet

In its current form, this program uses a modified form of the database from the Open Multilingual Wordnet. See Bond and Paik (2012) for details of its creation.

The structure of WordNet is organised around synsets which are linked to one another by semantic relations like hyponymy, meronymy, etc. Synsets are sets of synonyms which have approximately the same meaning, otherwise known as sense in WordNet.

The Open Multilingual Wordnet currently features 117,659 synsets, many of which are specialised vocabulary unfamiliar even to most native speakers. To prevent such words from being tested in the crossword puzzles and to concentrate on the set of words more useful to a beginner, the synsets used in automatic puzzle generation were limited to the 5000 core synsets in the original Princeton WordNet (Fellbaum, 1998), which are meant to correspond to the most frequently used word senses. However, these are based on the most frequent words and phrases in the British National Corpus (Burnard, 1995), which may result in a bias towards the concepts that appear most frequently in English and may not necessarily be indicative of the most frequent words in other languages due to cultural differences. See Boyd-Graber, Fellbaum, Osherson, and Schapire (2006) for more details on how the list of core synsets was compiled. Another reason why we had to limit the number of synsets is that the randomisation process involved in choosing items from a large set of words significantly slows the program down, and we needed it to be random to ensure that the automatically generated puzzles would be different for each run. However, the custom word lists make use of the full WordNet.

As the Open Multilingual Wordnet focuses only on adding lemmas at the moment, it lacks other information for alternative crossword clues. To remedy this, synset examples from the English and Japanese WordNets were copied over, along with definitions from the Japanese WordNet (Isahara, Bond, Uchimoto, Utiyama, & Kanzaki, 2008).

As a single synset typically features multiple synonyms, some of which

may be fairly obscure, we added frequency data collated from the NTU Multilingual Corpus (Tan & Bond, 2012) to the English and Japanese words. The lemmas of each synset are then sorted by frequency, with the most frequent lemmas being chosen over the more obscure ones to be used as puzzle solutions.

While importing word frequencies from external lists and corpora for all the other languages is outside the scope of this project, we came up with a heuristic to approximate the frequencies for all the words regardless of the language, using only information already available within WordNet. More commonly used words tend to have more meanings, either because having more meanings results in more contexts in which the word can be used, or because more common words are likely to have their meanings extended. For instance, the word 'dog' has 8 senses, one of which is to chase something, an action which is commonly associated with dogs. On the other hand, rare words like 'dormie', 'mudder' and 'pelecypodous' are specialised words with very precise meaning, hence they have only 1 sense. Hence we use the count of how many senses a word has as an approximation of its frequency. In a similar manner, to get the frequency of a synset, we count the number of languages in which it has words, on the assumption that the more common a concept is, the more languages would lexicalise it and include it in their WordNets. Although these methods do not obtain perfect results as there can be common words with few meanings and rare words with many meanings, they can still serve as a rough guideline for ordering our words and synsets by frequency.

In the current Japanese WordNet, the readings of kanji are not included as part of the words. As the hiragana readings are to be used in place of kanji when the puzzle solutions are in Japanese, we needed this data in our version of WordNet. Takayuki Kuribayashi has been working on adding this data to the Japanese WordNet, and while it is still a work-in-progress, we are using his data (Kuroda, Kuribayashi, Bond, Kanzaki, & Isahara, 2011) for the crossword program. However, the readings of words written in katakana will

not be included, as katakana is often used to spell words borrowed from other languages, and Japanese learners may be more likely to want to concentrate on learning native words instead of borrowings which may be from their own first language (particularly applicable to English speakers). The same process can also be used to include hanyu pinyin readings for Chinese characters.

To speed up the process of querying the database, we had to create additional search indexes on the pronunciation field of the word table, the language field of the sense table, and the frequency field of the sense table.

2.2 Puzzle Generator

The source to the original Python Crossword Puzzle Generator was created by Bryan Helmig and released under the BSD 2-Clause License. (Helmig, 2010)

The puzzle generator section of the script was mostly left unchanged, save for implementing the suggestion made by Daniel Nögel on a comment on Helmig's blog post releasing the source code. The original code suggested possible co-ordinates for the next word to be placed by iterating over every box of the crossword grid for each of the letters in each word to find a match, but the new script uses a Python dictionary to keep a list of co-ordinates for each letter based on the words already placed, resulting in a speed improvement.

Original code snippet of definition of function suggest_coord(self, word):

```
glc = -1
for given_letter in word.word: # cycle through letters
    in word
        glc += 1
        rowc = 0
        for row in self.grid: # cycle through rows
                rowc += 1
                colc = 0
```

```
                for cell in row: # cycle through
                    letters in rows
                        colc += 1
                        if given_letter == cell: #
                            check match letter in word
                            to letters in row
```

Modified code snippet of definition of function suggest_coord(self, word):

```
for glc, given_letter in enumerate(word.word): # cycle
    through letters in word
        for (colc, rowc) in self.letters[given_letter
            ]:
                if given_letter == self.grid[rowc][
                    colc]:
```

Added code snippet to definition of function set_cell(self, col, row, value):

```
self.letters[value].append((col, row))
self.letters[value] = list(set(self.letters[value]))
```

The original code was written for English crosswords, so the script had to be modified to use UTF-8 encoding in order to support foreign characters. All instances where foreign characters might be in the output had to be encoded, and input possibly containing foreign characters had to be decoded.

As the original puzzle generator only resulted in puzzles meant for print media, it had to be connected to the solving interface to allow users to play on the computer. It achieves this by creating a new webpage with the JavaScript app specific to each puzzle and jumping automatically to the hyperlink.

The key change that had to be made to allow automatic formulation of word lists to be fed into the puzzle generator was to link it to WordNet which is stored as an SQLite database, performing search queries based on the parameters of a HTML form to formulate a word list for puzzle generation. The HTML form was added so that users could customise their crossword puzzle by choosing the size of the grid, the language of the solution, the

language of the clue and the type of clue. The languages for the solutions and clues can be chosen from any of the languages featured in the Open Multilingual Wordnet. The clues can be a list of synonyms, definitions, examples, kanji characters (with hiragana readings as the solutions when Japanese is selected), or images. This change involved about 136 new lines of code.

The word list to be fed into the puzzle generator can either be randomly drawn from items in WordNet, or user-defined in a custom word list. The point which distinguishes the custom word lists in our program from other online crossword generation tools is that we do not require the user to provide both solution and clue. All that users need to provide is the solution, and the clue will be drawn from WordNet. The solution may be provided in the form of a word, a synset, or a combination of both. The user may also provide their own clue if they wish. The custom word lists allow the user to focus on studying specific vocabulary of their choice, and can be used to learn set vocabulary lists like the ones for language tests. Including custom word lists involved about 254 new lines of code.

2.3 Solving Interface

The source to the original JavaScript Crossword Engine was created by Pavel Simakov and released under LGPL. (Simakov, 2009)

As the interface was originally intended to be used for solving English crossword puzzles, it did not allow foreign characters or even accented letters to be entered as input. Our program was adapted to allow any unicode character to be entered, by making the function `oyCrosswordPuzzle.prototype .isValidChar(c)` return true regardless of input. However, that caused problems in some browsers like Firefox which interpret directional keys, backspaces, tabs and delete keys as keypress events, unlike other browsers which treat them as keydown events. This results in strange characters being entered into the boxes when the user presses one of those keys. To fix this, we inserted the following conditional in the `if (this.isValidChar(c)){}` code

segment of the function `oyCrosswordPuzzle.prototype.handleKeyPress(x, y, e)`:

```
var probkeys = (keyCode >= 37 && keyCode <= 40) ||
   keyCode == 8 || keyCode == 9 || keyCode == 46;
if (!probkeys) {
        // original code block here
}
```

The above code ensures that the puzzle boxes will only be filled with text input when the key pressed was not one of the problematic keys mentioned above.

Also, after entering an English letter the original program automatically moved the cursor to the next box of the focused word in the crossword puzzle, but that function did not work with characters entered through IMEs (necessary when typing Japanese words on an English keyboard), making it more troublesome to solve puzzles with solutions in languages like Japanese. We managed to fix this problem in our version of the program by adding the following code (split into 2 functions to mirror how keydown and keypress events were handled in the original code for the sake of consistency):

```
target.onkeyup = function(e){
        return oThis.handleKeyUp(x, y, e);
}

oyCrosswordPuzzle.prototype.handleKeyUp = function(x,
   y, e){
        if (!e) {
                e = window.event;
        }
        var keyCode = (e.which) ? e.which : e.keyCode;
        var target = this.inputCache.getElement(x, y);
        if(target.value.length >= 1) {
                var nonimekey = (keyCode >= 65 &&
```

```
                    keyCode <= 90) || (keyCode >= 48 &&
                     keyCode <= 57) || (keyCode >= 37
                    && keyCode <= 40) || keyCode == 8
                    || keyCode == 46 || keyCode == 16
                    || keyCode == 20 || keyCode == 222
                    || keyCode == 192;
                if (!nonimekey) {
                        this.moveToNextCell(x, y);
                }
        }
        return true;
}
```

The above code works by checking to see if the currently focused box is empty upon a keyup event. If it finds that the box is not empty, it moves the cursor to the next box automatically. However, this fix only works in recent versions of some modern browsers like Mozilla Firefox and Opera, but not in other browsers like Google Chrome due to the browsers handling JavaScript key events differently. Pressing the enter key to confirm a character entered through the Microsoft Japanese IME fires both keydown and keyup events in Opera and Internet Explorer, but only the keyup event in Firefox and only the keydown event in Chrome. As the check is triggered by the keyup event, this function fails in Chrome.

There was a bug in the original program where the last letter of the solutions could not be easily deleted by pressing the 'backspace' key unlike the rest of the letters in the word. We fixed this by changing the way 'backspace' keydown events were handled in the script, adding the following code before `this.moveToPrevCell(x, y);`:

```
e.preventDefault();
this.inputCache.getElement(x, y).value = ""
```

We also enabled other hotkeys like 'enter' and 'delete' to make the interface more intuitive and user-friendly. In addition, we inserted the following

code in the segment to be executed when ‘tab’ is pressed in order to toggle between horizontal and vertical directions when the focused box is part of two different solutions:

```
e.preventDefault();
if (this.dir == 0) {
        this.dir = 1;
} else {
        this.dir = 0;
}
this.unfocusOldCell();
this.focusNewCell(x, y, true);
break;
```

The only hint available for the original program was the option to reveal the entire word. For players who may be stuck on a particular solution but do not wish to give up altogether, we decided to introduce more hint options. We gave them the option to reveal a single letter in any position, and also the option to see an anagram of the entire word, and adjusted the scoring accordingly to reflect their use of the new hints. Alternative clues are also available when the player clicks on the number next to the original clue. These changes involved about 157 new lines of code, as well as many minor edits to the code in different parts of the script.

Entering input in languages which do not use the Latin alphabet or which involve accented characters may be difficult on standard keyboards when it is difficult to install the corresponding IME. Also, some tablet users may find that the virtual keyboard does not consistently appear when trying to solve the crossword. To fix this problem, we added an on-screen keyboard. Using this keyboard also solves the problem where the cursor does not automatically jump to the next box after entering characters using an IME in browsers like Google Chrome. Currently only English and Japanese hiragana is supported, though the code can be extended to account for other languages. Each button of the keyboard is created with markup language

like `<input type='button' class='scrnkey' onclick='scrnkey(value)' value='a'>`, and the function 'scrnkey' is defined as follows:

```
function scrnkey(item) {
        var x = oygCrosswordPuzzle.xpos;
        var y = oygCrosswordPuzzle.ypos;
        var target = oygCrosswordPuzzle.inputCache.
           getElement(x,y);
        if (!target.readOnly){
              target.value = item;
              oygCrosswordPuzzle.moveToNextCell(x, y
                 );
        }
}
```

In order to allow the user to learn more about unfamiliar words, we also added some code to show hyperlinks to the relevant synset in WordNet for each of the items in the crossword after the puzzle is solved. To that end, we needed to add 'synset' as a new attribute of the 'word' class in the puzzle generator program.

2.4 User System

The login sessions are implemented in PHP, using a table of users in a MySQL database. Having a user system allows the program to remember the custom settings for the last generated puzzle, let users use custom word lists and record user scores. It also allows for the possibility of remembering clue-solution pairs that the user had trouble with (as they used hints), which are features that may be useful to implement in future.

3 Demo

Figure 1 shows the form by which the user customises the crossword and generates the puzzle by clicking on the button.

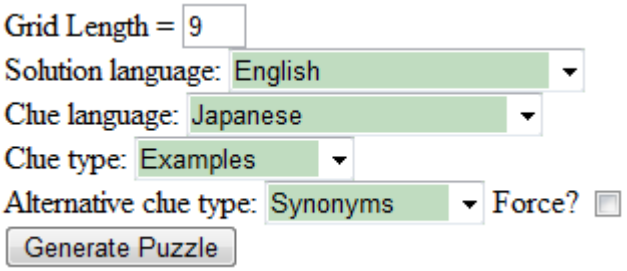

Figure 1: Options form for puzzle in Figure 6

Theoretically the grid can be of any size, but for practical reasons the program limits the width in boxes to numbers from 5 to 50. A 5x5 grid can accommodate about 5 clues, while a 50x50 grid had 357 clues when tested, though the numbers will vary with different options chosen. The 50x50 grid is too large to fit on most monitors, and a 13x13 grid should be comfortable for most users with about 20 clues when the solution is Japanese and 12 clues when the solution is English, since English words tend to be longer than Japanese and hence harder to place on a crossword grid.

To reiterate, the languages are all those available in the Open Multilingual Wordnet, and the clue types are a list of synonyms, definitions, examples, kanji characters (with hiragana solutions) and images. Alternative clues may be viewed by clicking on the number in the list of clues in the solving interface. Ticking the 'Force?' checkbox ensures that there will be alternative clues for every item, but doing so filters out items for which no alternative clue is available, which may drastically limit the words that may appear in the solution if the alternative clue type chosen is one for which not many words have data.

As the current database only has definitions and examples for English, Japanese and Albanian, if the user selects definitions or examples for any

other language, they will be given synonyms as clues instead. If kanji is selected as the clue type, the crossword generator will automatically treat Japanese as the language of both solution and clue.

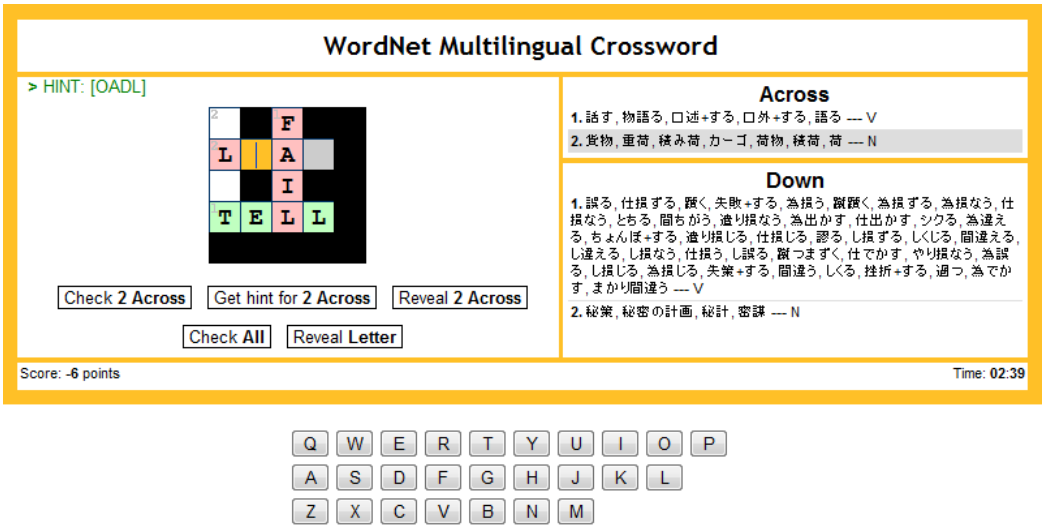

Figure 2: What the user sees while playing

Figure 2 shows what the user sees while solving the crossword puzzle. Empty boxes are white, and remain white even when text is entered, only changing colour when the solution is either revealed or checked and found to be correct. In this case, the user has revealed the word 'fail' and the letter 'L' by using the hint buttons below the grid, and the results are marked in pink boxes. The user has also correctly guessed the word 'tell', which was marked in green boxes when the user clicked on either the corresponding 'check word' or 'check all' button. The anagram hint provided above the puzzle appeared when the user clicked the corresponding 'get hint' button. Everytime a hint or check button is clicked, the score at the bottom is updated.

The on-screen keyboard is below the solving interface. Since the solutions are in English, the keyboard only shows the English alphabet. When the solution language is Japanese, the hiragana keyboard in Figure 3 will appear.

After the whole puzzle is solved, a list of the solution words will appear below the crossword, each containing a hyperlink to the corresponding synset

Figure 3: On-screen keyboard for Japanese hiragana

in the Open Multilingual Wordnet. This allows the user to learn more about an unfamiliar word by viewing all the information that WordNet has for that synset, including synonyms, definitions, examples and semantic links with other concepts. Also, at this point the full scores are automatically sent to the server and logged for future reference, including data like the puzzle's unique identification number, time taken to solve the puzzle, number of times each hint or check button was clicked, points gained for correctly guessing solutions, points lost for asking for hints and overall score.

The word list used to generate the crossword is by default drawn randomly from WordNet. However, if the user provides a custom word list by entering it in the correct format in a text box on the site and clicking the 'save' button, that list would be used instead. The easiest way to format the word list would be to use the 'w' tag and provide only the solution word. The word can be in any of the supported languages as long as the correct solution language is selected in the options. In the case of Japanese, text can be entered as either hiragana or kanji, but the solutions will still be in hiragana form. The 'w' tag can easily be used to input standard vocabulary lists like the one for the Japanese Language Proficiency Test (JLPT).

Figure 4 shows a puzzle generated with the JLPT N5 word list written in hiragana. Even though definitions were selected as the clue, some examples appear as well as the English WordNet includes the examples in the definition. As an example of how such custom word lists should look, the first 5 lines are as follows:

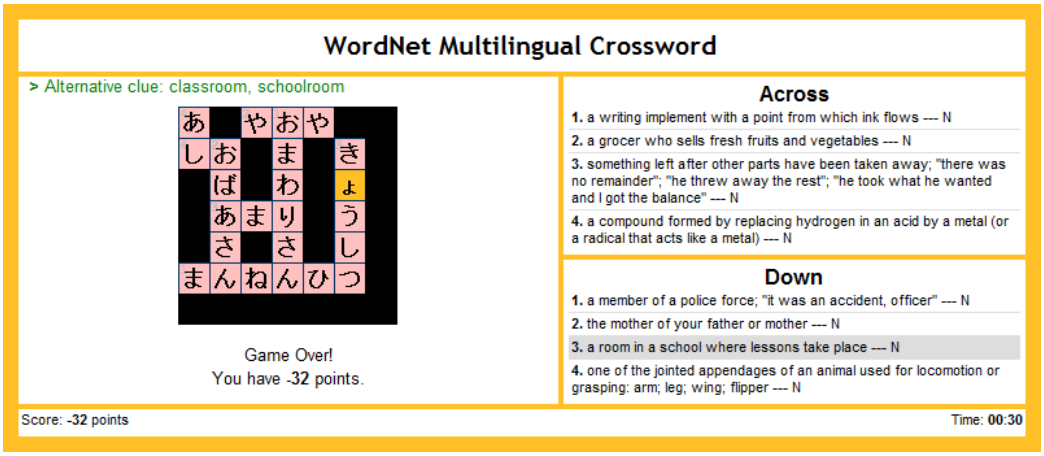

Figure 4: Crossword Example: 7x7, Japanese solution, English definition as clue with synonyms as alternative, custom word list in hiragana

w あう
w あお
w あおい
w あか
w あかい

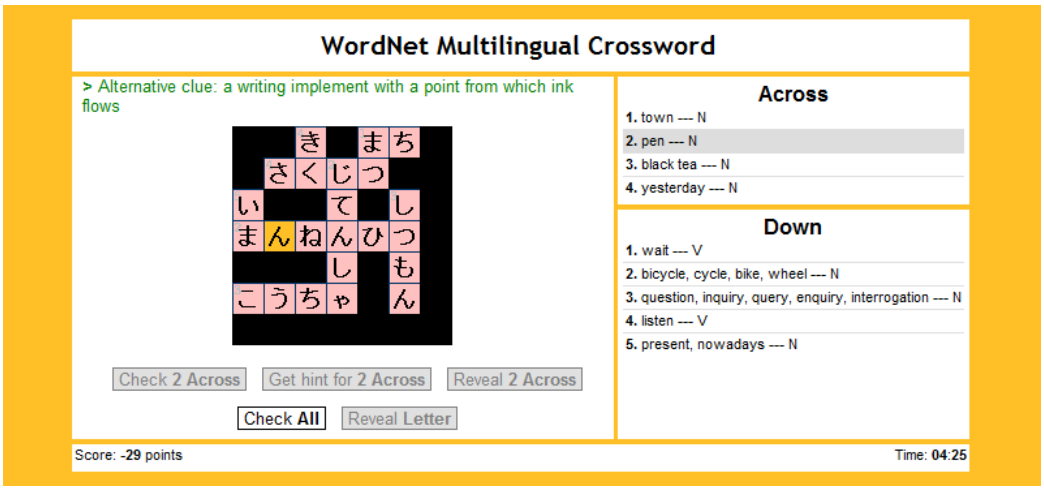

Figure 5: Crossword Example: 7x7, Japanese solution, English synonyms as clue with definition as alternative, custom word list in kanji

Figure 5 shows a puzzle generated with the JLPT N5 word list written in kanji, producing results similar to the hiragana version as intended. The first 5 lines of the custom word list are as follows:

```
w 会う
w 青
w 青い
w 赤
w 赤い
```

Instead of providing a word as the solution, users can choose to provide only the synset instead by using the 's' tag. Alternatively, users can provide both synset and word with the 'sw' tag, with everything separated with a single space. The user can also define both the solution and the clue with the tag 'wc', with the alternative clue as an optional parameter following the first clue separated by a '|'. The different tags can be combined in the same word list. As an example, the following word list produces the puzzle in Figure 6:

```
wc earth the planet on which we live --- N
wc sun great ball of gas --- N|in the sky during the day
s 04379243-n
sw 09444100-n star
```

If the text box for the custom word list is empty, the solutions and clues will all be drawn randomly from WordNet. Figures 7 to 9 are examples of custom puzzles generated with different settings, with the details provided in the captions below the figures. The alternative clues shown above the grid correspond to the highlighted clue.

When images are selected as the clue, the list of images are displayed as small icons, but bigger images pop up at the top right-hand-corner when the mouse hovers over the icon.

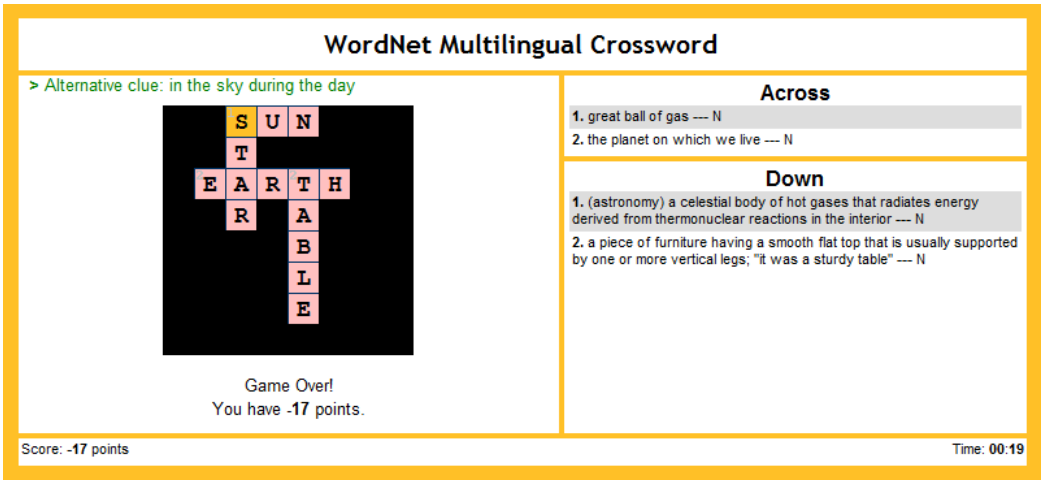

Figure 6: Crossword Example: 8x8, English solution, English definitions as clue with examples as alternative, custom word list

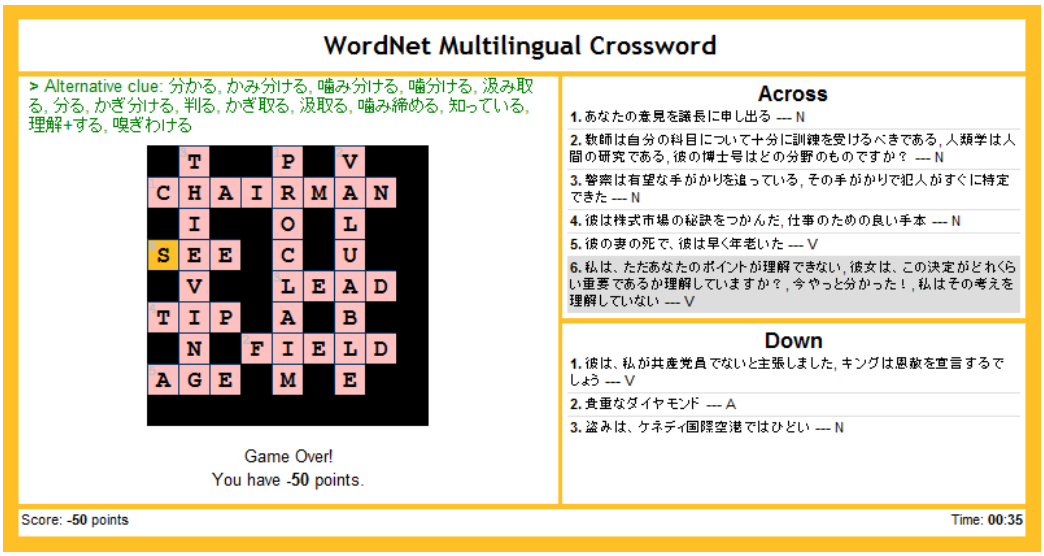

Figure 7: Crossword Example: 9x9, English solution, Japanese examples as clue with synonyms as alternative, core synsets

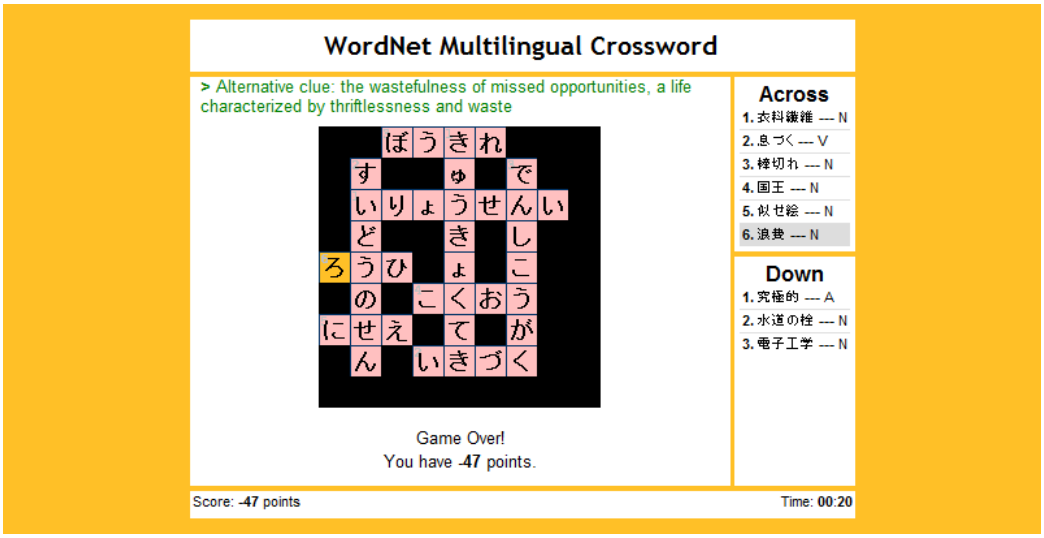

Figure 8: Crossword Example: 9x9, Japanese hiragana readings as solution, kanji as clue with English examples as alternative, core synsets

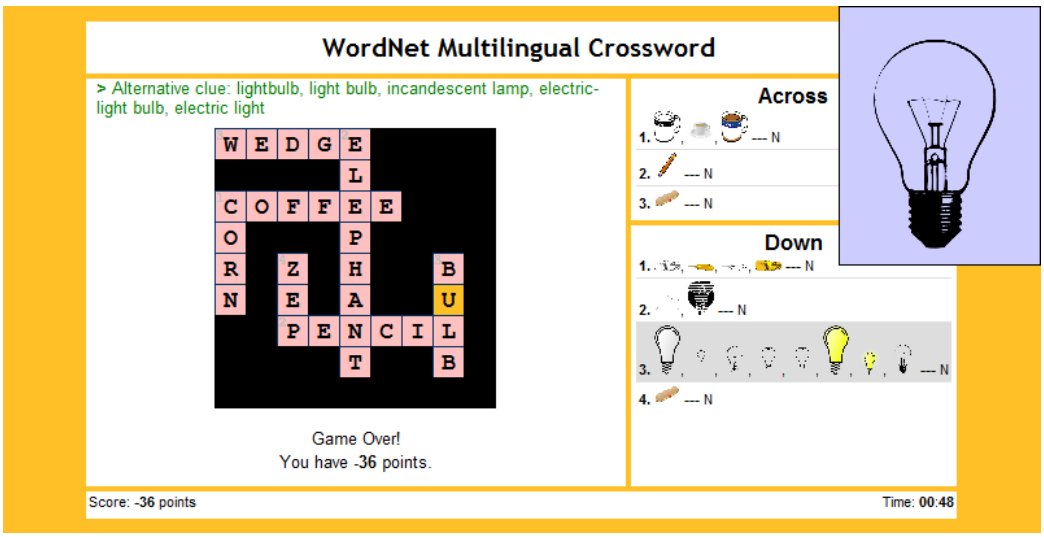

Figure 9: Crossword Example: 9x9, English solution, images as clue with English synonyms as alternative, core synsets

4 Methodology

4.1 Participants

The participants are 38 students taking beginner-level Japanese courses in NTU, with 27 from Level 1 and 11 from Level 2. Participants were restricted to only 2 levels of Japanese students as custom vocabulary lists and tests had to be created for each level and language. The beginner levels were selected as they have the highest enrolment, ensuring a wide pool of eligible participants, and also because the multilingual crosswords are primarily targeted at beginners. Participants were recruited by uploading an announcement calling for volunteers to register for the study on the course e-learning website, and all those from Levels 1 and 2 who signed up were accepted provided they were available for the study timeslots.

They were randomly assigned to 6 conditions describing the category of crossword puzzle they were supposed to solve. The conditions are described in Table 1.

Group	Clue	Solution
A	English definition	Japanese word in hiragana
B	English word	Japanese word in hiragana
C	Japanese word in kanji	Hiragana reading
D	Japanese word in kanji	English word
E	Japanese word in hiragana	English word
F	Control: No crossword puzzles	

Table 1: Experimental groups to which participants were assigned

4.2 Procedure

Participants were intended to learn a 25-item vocabulary list through doing crossword puzzles in this study. Since participants were not supposed to be already familiar with the vocabulary, words included in the school's vo-

cabulary list for their level were not used in the study. Words were selected from the vocabulary lists from the basic levels of the Japanese Language Proficiency Test (JLPT) to ensure that they were appropriate for the students' level. Level 1 students were to learn words at the JLPT N5 level while Level 2 students were given JLPT N4 words, with the word lists being taken from Waller (2010). All words chosen for the study included at least one Kanji character so we could examine more aspects of the participants' learning of Japanese vocabulary. To measure how much the participants learnt, we gave them a vocabulary test once at the start of the study and again at the end and calculated the improvement in their scores.

In Week 1 of the study, participants were asked to read and sign a consent form explaining the relevant details of the study before beginning any activities. They were then given a pre-activity vocabulary test and allowed up to 30 minutes to complete it, following which they were asked to look through a review sheet (covering the spellings, pronunciations, translations and definitions of all 25 words) for 3 minutes. All participants in the same Japanese level did the same test, which consisted of 25 questions each corresponding to one Japanese word in the vocabulary list they were intended to learn during the study. The questions were separated into 5 sections (5 questions each) corresponding to the categories of crosswords as outlined in Table 1, testing their knowledge of a different aspect of vocabulary (e.g. spelling, pronunciation, translation and definition). Example questions with the answer key underlined are as follows:

a. Give the Japanese word that fits the English definition below in hiragana:

    - "Having a temperature slightly higher than usual, but still pleasant; a mild temperature." - <u>あたたかい</u>

b. Give the Japanese translation of the English word below in hiragana:

    - warm - <u>あたたかい</u>

c. Give the readings of the Japanese words below in hiragana:

   - 暖かい - あたたかい

d. Give the simple English translation of the Japanese word below written in kanji:

   - 暖かい - warm

e. Give the simple English translation of the Japanese word below written in hiragana:

   - あたたかい - warm

In Week 2, they were asked to do 2 crossword puzzles online at their own convenience. Then in Week 3, they were asked to do 1 crossword puzzle. No time limits were set and they were allowed to use as many hints as they liked as they were not expected to know all the answers. The program did not allow them to do more than the allotted number of puzzles per week, and records were kept on the server to ensure they met their target for the week. The timings of the puzzles were flexible and only 3 puzzles were set as students are likely to be busy with their coursework and requiring too much of them would likely result in their dropping out of the study.

For the purposes of this experiment, a version of the crossword program with limited functions was used. Since the experiment was intended to test how the crossword program could help students to remember new vocabulary, only the 25-item vocabulary list was used to generate the crosswords, without allowing users the option of defining their own lists or randomly picking items from WordNet. Users were also not allowed to customise options like the size of the crossword grid, and the languages of the solution and clues and the type of clues were set according to their experimental group.

In Week 3 after all 3 crossword puzzles were completed, participants were given a post-activity vocabulary test, with the same 25 words but with their corresponding questions placed in a different section to measure their knowledge of a different aspect of vocabulary. Only one aspect of each word was

tested in a single test, to keep the test short and to ensure that participants do not learn the correct answers from the questions themselves.

To ensure consistency in grading, I graded all the vocabulary tests by myself. Any reasonable answer that fit the wording of the question was given the full mark, even if it did not match the answer key exactly. If the part-of-speech of the answer was not clear from the question, alternative parts-of-speech were allowed. Both polite and plain forms were accepted for Japanese words. Half a mark was given for answers with spelling mistakes. If the participant did not give what the question asked for, for instance by providing the hiragana reading to the kanji when the question asked for the English translation, no mark was given as each section only measures one specific aspect of vocabulary knowledge and other aspects are deemed irrelevant.

5 Results and Discussion

Tables 2 and 3 list the descriptive statistics for the total test scores, aggregated by level and group. 'Pre' refers to the pre-activity test, 'pos' to the post-activity test, 'dif' the absolute difference between the two, 'dip' the relative difference and 'T' stands for total score of all the sections in the test. Positive 'dif' and 'dip' values indicate improvement while negative values mean that the score worsened. While 'dif' is the difference in the actual marks (PosT - PreT), 'dip' is the difference in scores as a percentage of the potential for improvement, or ((PosT - PreT)/(25 - PreT))x100%, with 25 being full score. The reason for showing relative improvement is that if a partipant scores 4 marks out of 5 in a section during the pre-activity test and then 5 marks in the post-activity test, the improvement of 1 mark is not slight as the scores have reached the ceiling and a greater improvement might have been possible had there been more questions on the test. There were about 5 participants per group for Level 1 and 2 for Level 2, but 4 participants dropped out of the experiment.

Level	Group		N	Minimum	Maximum	Mean	Std. Deviation
1	A (def-hira)	PreT	5	3.5	11.0	7.200	3.5107
		PosT	5	12.5	22.0	16.900	4.0062
		DifT	5	5.5	16.5	9.700	4.3099
		DipT	5	.3793	.7857	.545220	.2114697
		Valid N (listwise)	5				
	B (eng-hira)	PreT	4	.5	8.5	6.000	3.6968
		PosT	4	1.0	17.0	10.250	6.7392
		DifT	4	.5	8.5	4.250	3.3789
		DipT	4	.0204	.5152	.248175	.2085556
		Valid N (listwise)	4				
	C (kan-hira)	PreT	4	2.0	14.0	6.750	5.3619
		PosT	4	12.5	19.0	14.375	3.0923
		DifT	4	5.0	11.0	7.625	3.0923
		DipT	4	.2857	.4783	.415100	.0875770
		Valid N (listwise)	4				
	D (kan-eng)	PreT	5	.0	20.5	6.500	8.1777
		PosT	5	5.0	24.0	10.300	7.7427
		DifT	5	1.0	7.5	3.800	2.3345
		DipT	5	.0526	.7778	.290000	.2865961
		Valid N (listwise)	5				
	E (hira-eng)	PreT	5	4.5	17.0	10.200	4.9320
		PosT	5	6.0	22.0	11.800	6.6389
		DifT	5	-.5	5.0	1.600	2.0736
		DipT	5	-.0278	.6250	.165540	.2628721
		Valid N (listwise)	5				
	F (control)	PreT	4	5.0	15.5	10.000	4.8477
		PosT	4	5.5	19.0	12.875	5.6917
		DifT	4	.5	5.0	2.875	1.8875
		DipT	4	.0250	.3684	.217800	.1457956
		Valid N (listwise)	4				

Table 2: Descriptive statistics for Level 1

Level	Group		N	Minimum	Maximum	Mean	Std. Deviation
2	A	PreT	2	2.0	11.5	6.750	6.7175
		PosT	2	7.0	14.0	10.500	4.9497
	(def-hira)	DifT	2	2.5	5.0	3.750	1.7678
		DipT	2	.1852	.2174	.201300	.0227688
		Valid N (listwise)	2				
	B	PreT	2	8.5	8.5	8.500	.0000
		PosT	2	12.5	19.5	16.000	4.9497
	(eng-hira)	DifT	2	4.0	11.0	7.500	4.9497
		DipT	2	.2424	.6667	.454550	.3000254
		Valid N (listwise)	2				
	C	PreT	1	6.0	6.0	6.000	.
		PosT	1	18.0	18.0	18.000	.
	(kan-hira)	DifT	1	12.0	12.0	12.000	.
		DipT	1	.6316	.6316	.631600	.
		Valid N (listwise)	1				
	D	PreT	2	8.5	9.0	8.750	.3536
		PosT	2	14.5	15.0	14.750	.3536
	(kan-eng)	DifT	2	5.5	6.5	6.000	.7071
		DipT	2	.3438	.3939	.368850	.0354260
		Valid N (listwise)	2				
	E	PreT	2	.0	3.5	1.750	2.4749
		PosT	2	10.0	11.0	10.500	.7071
	(hira-eng)	DifT	2	7.5	10.0	8.750	1.7678
		DipT	2	.3488	.4000	.374400	.0362039
		Valid N (listwise)	2				
	F	PreT	2	4.0	6.0	5.000	1.4142
		PosT	2	11.0	11.0	11.000	.0000
	(control)	DifT	2	5.0	7.0	6.000	1.4142
		DipT	2	.2632	.3333	.298250	.0495682
		Valid N (listwise)	2				

Table 3: Descriptive statistics for Level 2

In Level 1, the mean pre-activity test scores for groups A to D were around 6 to 7, while groups E and F scored around 10 marks on average. This difference is just noise as the pre-activity experimental conditions were the same for all groups. The mean post-activity test scores were scattered from 10 to 17 marks. In Level 2, the mean pre-activity test scores for groups A to D were around 6 to 9, while the mean score for groups E and F was about 2 and 5 respectively. The mean post-activity test scores were scattered from 10 to 18 marks.

As our experiment is designed to test the improvement of vocabulary after doing the crossword puzzles, we will focus on examining the 'dif' and 'dip' scores instead of the 'pre' and 'post' by presenting them in graphical form. We found that the patterns for Level 1 and 2 were quite different as they took separate tests which may have been of different difficulty levels relative to their current language proficiency, hence their data cannot fairly be merged. As there are only 1 or 2 people in each Level 2 group, we believe that data set by itself is highly subject to noise and those results are not significant, so we will only discuss the results for Level 1.

For figures 10 to 13, the capital letters A to F represent the experimental groups as described in Table 1 with F being the control group, and the small letters a to e represent the sections in the vocabulary test as described in Section 4.2. The same letters are used as they correspond to the same combination of solution/clue and question/answer pairs, with Group A being assigned crossword puzzles designed to directly improve their performance on section a. Figures 10 and 11 are broken down by test section while figures 12 and 13 show total scores; and figures 10 and 12 show absolute improvement while figures 11 and 13 show relative.

Contrary to expectations, neither of the graphs broken down by test section show any particular trend that would suggest that the type of crossword puzzle directly affects the scores of the corresponding test section except for Group D on section d and possibly E on e, though the patterns for absolute and relative improvement are largely similar. Due to the small sample size,

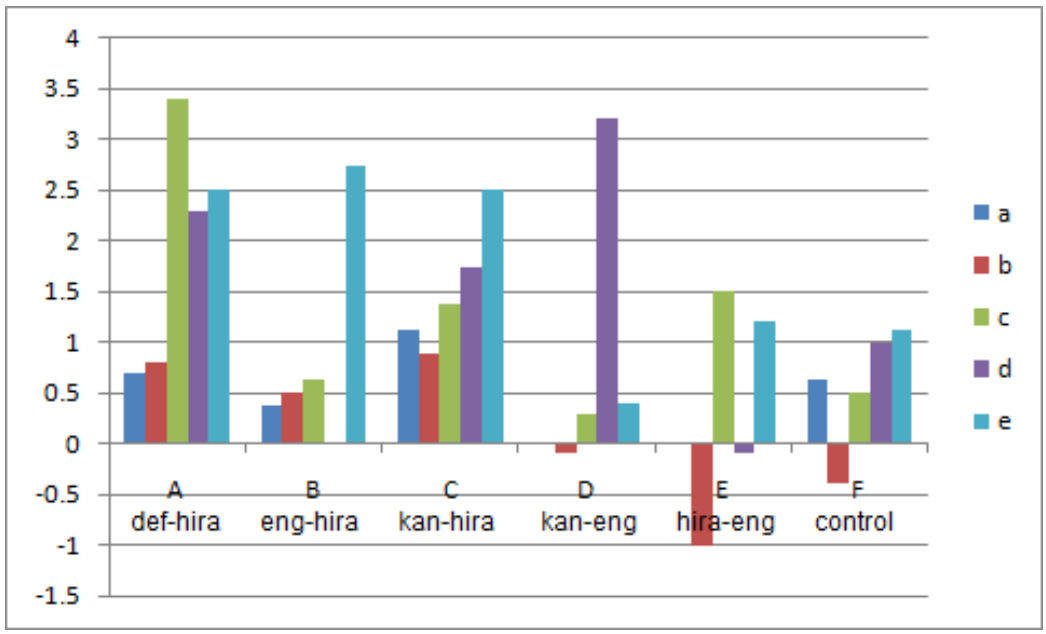

Figure 10: Bar graph for absolute improvement by test section in Level 1

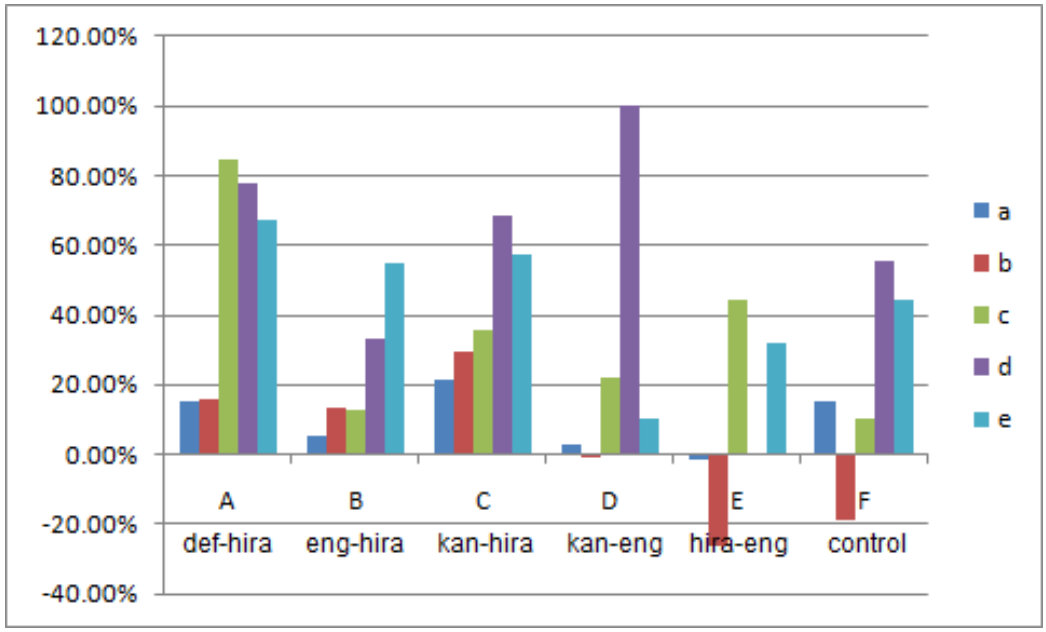

Figure 11: Bar graph for relative improvement by test section in Level 1

it is unclear if this lack of a trend is because there is no relationship between puzzle type and section score or due to the pattern being obscured by noise.

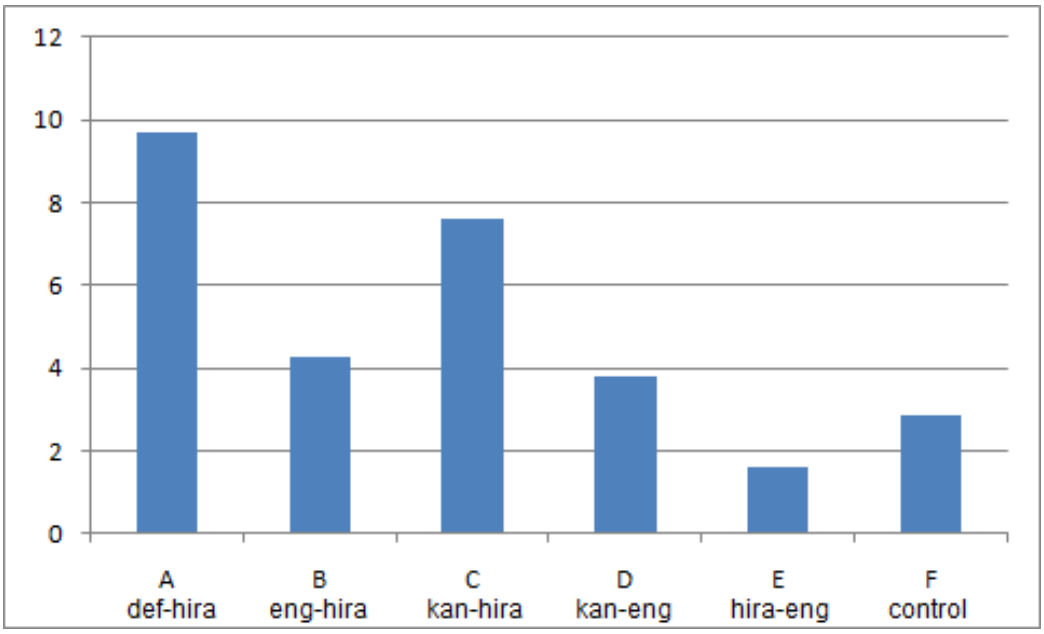

Figure 12: Bar graph for total absolute improvement in Level 1

In general, all experimental groups except for Group E show more overall improvement than the control group F, with a largely similar trend for both absolute and relative improvement, indicating that the crossword puzzles did indeed help participants to learn vocabulary. Again, due to the small sample size it is not clear if puzzle type E is actually ineffective in improving vocabulary or if these results were due to chance. In general, groups A and C improved the most, followed by B and D which show slightly greater improvement than the control, while E improved less than the control. The relative effectiveness of the puzzle types A and C might be explained by the fact that the solutions were in Japanese and hence solving those puzzles forced participants to think in Japanese thereby making them more effective in learning vocabulary, and it may be due to noise that the effect did not transfer to type B which also had Japanese solutions. In contrast, the solutions for D and E are in English which might require less cognitive effort to recall (reading the

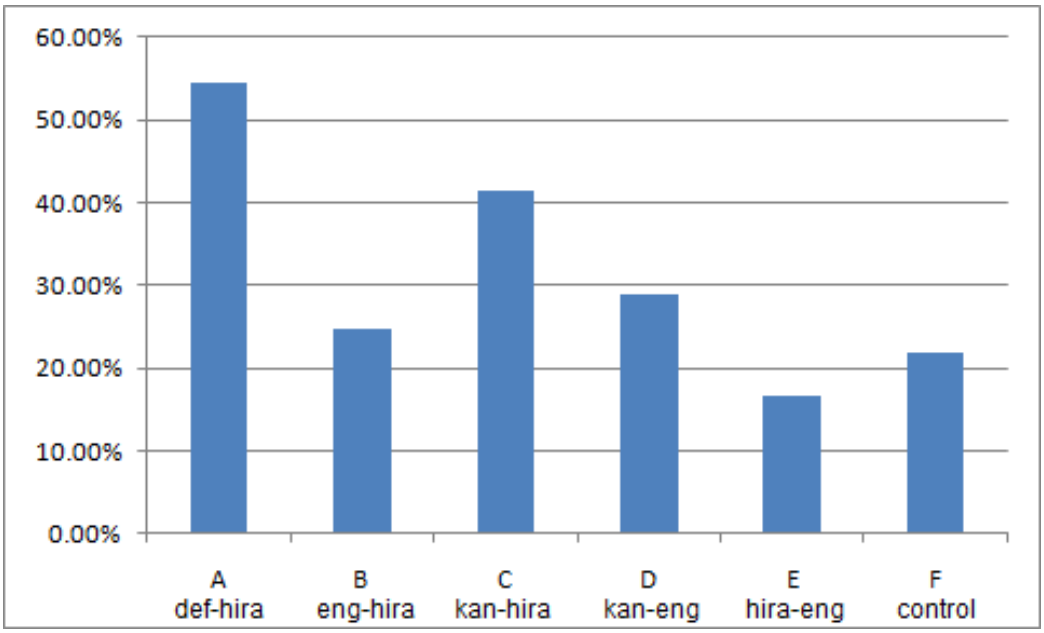

Figure 13: Bar graph for total relative improvement in Level 1

Japanese clue only requires them to recognise, not recall, the words), hence the association between solution and clue might be encoded less deeply in the memory.

The lack of a definite pattern may also stem from too few crossword puzzles being assigned rather than point to the inefficacy of crossword puzzles in aiding vocabulary acquisition. One participant volunteered feedback to the effect that the number of puzzles might have been too small, and she felt that she would have improved more had she done 3 puzzles a week instead. It is true that assigning more puzzles is likely to lead to more conclusive results, but we decided against it for this experiment because of the risk of losing participants if the required commitment level were too high.

6 Conclusion

Due to the small sample size the results of this experiment are only indicative. To further evaluate the program and the effects of different types of

crossword clues on vocabulary acquisition, future experiments could be run with a larger pool of participants over a longer time period with a greater frequency of crossword puzzles per week.

In addition, there are many features outside the scope of this study that could prove useful if introduced in the program, but were unfortunately too time-consuming to implement during the course of this project.

To begin with, future versions of the program could improve the experience for users who prefer a monolingual puzzle, with the same language for both the solutions and the clues. Currently the program removes the solution word from the list of synonyms in the clue, but it does not account for alternative spellings of the same word. In future, the edit distance between two words in the same synset may be calculated to determine if they should be treated as the same word spelt differently, and alternative spellings of the solution word can be removed from the list of synonyms. Also, the lemmas should be removed from the example sentences.

We could make use of the encoded semantic relations between synsets in WordNet to provide more types of clues like hypernyms, hyponyms, antonyms, meronyms, etc. where such data is available. In addition, the hypernyms and hyponyms can also be used to organise vocabulary lists around a common theme.

Though our study focuses on WordNet as the only word database used in the program, we could enrich our database with information from other sources. Word frequencies can be obtained through more accurate frequency lists specific to each language and based on larger corpora. Also, concordance, collocation and n-gram information would allow users to study in greater detail how each word is used in context.

While the current project focuses on crossword puzzles in English and Japanese, the Open Multilingual WordNet features many other languages. To provide more support for these other languages, we could add the respective synset examples and definitions to WordNet to make more alternative crossword clues available, and also provide on-screen keyboards specific to

those other languages.

The solving interface does not account for solutions in languages which are written right-to-left like Arabic, Persian and Hebrew. Though we can easily reverse the characters in each solution word to make them display correctly, typing them into the interface would still be difficult as the cursor would move in the wrong direction. Furthermore, it is unclear how the words should be represented when the solution is vertical.

Users' responses to specific solution-clue pairs may be recorded, particularly on custom word lists. As asking for hints indicates some trouble with solving that item, those items could be scheduled to appear again on subsequent puzzles, while items solved without any hints could be made less likely to reappear.

The game could be made more fun by allowing competition on a few shared puzzles. As player scores are already recorded, they could be added to a leaderboard for players to compare their performance with others.

Lastly, while the current website makes use of much screen space, a mobile version could be made to allow smartphone users to solve crossword puzzles on the go.

A Vocabulary Test Questions with Model Answers

Pre-activity vocabulary test for Level 1

Participant ID: _____

a. Give the Japanese word that fits the English definition below in hiragana:
e.g. the season when the leaves fall from the trees - あき

1. a white crystalline carbohydrate used as a sweetener and preservative - さとう
2. of the color intermediate between green and orange in the color spectrum; of something resembling the color of an egg yolk - きいろい
3. a pocket-size case for holding papers and paper money - さいふ
4. not easy; requiring great physical or mental effort to accomplish or comprehend or endure - むずかしい
5. something that provides access (to get in) - いりぐち

b. Give the Japanese translation of the English word below in hiragana:
e.g. autumn - あき

1. head - あたま
2. time - じかん
3. small - ちいさい
4. animal - どうぶつ
5. electricity - でんき

c. Give the readings of the Japanese words below in hiragana:
e.g. 秋- あき

1. 暖かい - あたたかい
2. 仕事 - しごと
3. 問題 - もんだい
4. 料理 - りょうり
5. 質問 -しつもん

d. Give the simple English translation of the Japanese word below written in kanji:
e.g. 秋 - autumn

1. 疲れました - tired
2. 練習 - practice
3. 遊びます - play
4. 全部 - whole/all
5. 忘れます - forget

e. Give the simple English translation of the Japanese word below written in hiragana:
e.g. あき - autumn

1. てんき - weather
2. つよい - strong
3. あたらしい - new
4. おおきい - big
5. つめたい - cold

Post-activity vocabulary test for Level 1

Participant ID: _____

a. Give the Japanese word that fits the English definition below in hiragana:
e.g. the season when the leaves fall from the trees - あき

1. having a low or inadequate temperature (not the weather) - つめたい
2. dismiss from the mind; stop remembering - わすれます
3. limited or below average in number or quantity or magnitude or extent - ちいさい
4. systematic training by multiple repetitions - れんしゅう
5. the principal activity in your life that you do to earn money - しごと

b. Give the Japanese translation of the English word below in hiragana:
e.g. autumn - あき

1. cooking (noun) - りょうり
2. big - おおきい
3. entrance - いりぐち
4. to play - あそびます
5. sugar - さとう

c. Give the readings of the Japanese words below in hiragana:
e.g. 秋- あき

1. 時間 - じかん
2. 強い - つよい
3. 疲れました - つかれました
4. 新しい - あたらしい
5. 天気 - てんき

d. Give the simple English translation of the Japanese word below written in kanji:
e.g. 秋 - autumn

1. 質問 - question
2. 難しい - difficult
3. 財布 - wallet
4. 頭 - head
5. 電気 - electricity

e. Give the simple English translation of the Japanese word below written in hiragana:
e.g. あき - autumn

1. きいろい - yellow
2. どうぶつ - animal
3. もんだい - question/problem
4. あたたかい - warm
5. ぜんぶ - whole/all

Pre-activity vocabulary test for Level 2

Participant ID: _____

a. Give the Japanese word that fits the English definition below in hiragana:
e.g. the season when the leaves fall from the trees - あき

1. to assess the similarities and differences between two or more things - くらべます
2. the way in which the words of a language are made to sound when speaking - はつおん
3. having a feeling of shameful discomfort - はずかしい
4. an unfortunate mishap; especially one causing damage or injury - じこ
5. to give money or other compensation to in exchange for goods or services - はらいます

b. Give the Japanese translation of the English word below in hiragana:
e.g. autumn - あき

1. to dance - おどります
2. international - こくさい
3. freedom - じゆう
4. failure - しっぱい
5. reason - りゆう

c. Give the readings of the Japanese words below in hiragana:
e.g. 秋- あき

1. 厳しい - きびしい
2. 正しい - ただしい
3. 泥棒 - どろぼう
4. 会話 - かいわ
5. 意見 - いけん

d. Give the simple English translation of the Japanese word below written in kanji:
e.g. 秋 - autumn

1. 役に立ちます- useful/helpful
2. 続けます- continue
3. 優しい - kind
4. 具合 - health/condition
5. 手伝います - help

e. Give the simple English translation of the Japanese word below written in hiragana:
e.g. あき - autumn

1. とくべつ - special
2. しゃかい - society
3. げんいん - cause
4. ひるま - daytime
5. にげます - escape

Post-activity vocabulary test for Level 2

Participant ID: _____

a. Give the Japanese word that fits the English definition below in hiragana:
e.g. the season when the leaves fall from the trees - あき

1. lack of success - しっぱい
2. to have a practical or beneficial use - やくにたちます
3. one who has carried out a theft - どろぼう
4. to provide assistance to (someone or something) - てつだいます
5. distinguished by a unique or unusual quality - とくべつ

b. Give the Japanese translation of the English word below in hiragana:
e.g. autumn - あき

1. to escape - にげます
2. to pay - はらいます
3. to continue - つづけます
4. opinion - いけん
5. to compare - くらべます

c. Give the readings of the Japanese words below in hiragana:
e.g. 秋- あき

1. 理由 - りゆう
2. 原因 - げんいん
3. 優しい - やさしい
4. 具合 - ぐあい
5. 社会 - しゃかい

d. Give the simple English translation of the Japanese word below written in kanji:
e.g. 秋 - autumn

1. 恥ずかしい - embarrassed
2. 踊ります- dance
3. 昼間 - daytime
4. 発音 - pronunciation
5. 厳しい - strict/harsh

e. Give the simple English translation of the Japanese word below written in hiragana:
e.g. あき - autumn

1. ただしい - correct
2. こくさい - international
3. じこ - accident
4. じゆう - freedom
5. かいわ - conversation

References

Aherne, A., & Vogel, C. (2006). Wordnet enhanced automatic crossword generation. In P. Sojka, K.-S. Choi, C. Fellbaum, & P. Vossen (Eds.), *Proceedings of the 3rd international wordnet conference* (p. 139-145).

Alimemaj, Z. (2010). Using crossword puzzles to teach grammar and improve communication strategies. In A. Shafaei & M. Nejati (Eds.), *Annals of language and learning: Proceedings of the 2009 international online language conference (iolc 2009)* (p. 298-303).

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, *4*(5), 316–321.

Bahrick, H. P., & Phelps, E. (1987). Retention of spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *13*(2), 344–349. doi: 10.1037/0278-7393.13.2.344

Bond, F., & Paik, K. (2012). A survey of wordnets and their licenses. In *Proceedings of the 6th global wordnet conference (gwc 2012)* (p. 64–71).

Boyd-Graber, J., Fellbaum, C., Osherson, D., & Schapire, R. (2006). Adding dense, weighted connections to wordnet. In *Proceedings of the third global wordnet meeting, jeju island, korea, january 2006.*

Burnard, L. (1995). *User reference guide for the British National Corpus* (Tech. Rep.). Oxford University Computing Services.

Elgort, I. (2007). *The role of intentional decontextualised learning in second language vocabulary acquisition: Evidence from primed lexical decision tasks with advanced bilinguals.* Victoria University of Wellington, Wellington, New Zealand. (Unpublished doctoral dissertation)

Fellbaum, C. (Ed.). (1998). *Wordnet: An electronic lexical database.* MIT Press.

Fitzpatrick, T., Al-Qarni, I., & Meara, P. (2008). Intensive vocabulary learning: A case study. *Language Learning Journal*, *36*(2), 239–248. doi: 10.1080/09571730802390759

Helmig, B. (2010). *Python crossword puzzle generator.* Retrieved from

http://bryanhelmig.com/python-crossword-puzzle-generator/

Isahara, H., Bond, F., Uchimoto, K., Utiyama, M., & Kanzaki, K. (2008). Development of japanese wordnet. In *Lrec-2008.*

Kuroda, K., Kuribayashi, T., Bond, F., Kanzaki, K., & Isahara, H. (2011). Orthographic variants and multilingual sense tagging with the Japanese WordNet. In *17th annual meeting of the association for natural language processing* (p. A4-1). Toyohashi.

Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: a critical investigation of flashcard software. *Computer Assisted Language Learning*, *24*(1), 17-38.

Nation, I. S. P. (1980). Strategies for receptive vocabulary learning. *RELC Guidelines*, *3*, 18–23.

Simakov, P. (2009). *Javascript crossword engine – source released lgpl.* Retrieved from http://www.softwaresecretweapons.com/jspwiki/cword/

Tan, L., & Bond, F. (2012). Building and annotating the linguistically diverse NTU-MC (NTU-multilingual corpus). *International Journal of Asian Language Processing*, *22*(4), 161–174.

Thorndike, E. L. (1908). Memory for paired associates. *Psychological Review*, *15*(2), 122–138. doi: 10.1037/h0073570

Waller, J. (2010). *Jlpt n5 vocab list.* Retrieved from http://www.tanos.co.uk/jlpt/jlpt5/vocab/

Webb, S. A. (2002). *Investigating the effects of learning tasks on vocabulary knowledge.* Victoria University of Wellington, Wellington, New Zealand. (Unpublished doctoral dissertation)

Webb, S. A. (2009). The effects of pre-learning vocabulary on reading comprehension and writing. *Canadian Modern Language Review*, *65*(3), 441–470. doi: 10.1353/cml.0.0046