phloat : Integrated Writing Environment for ESL learners

$Yuta Hayashibe^1$ $Masato Hagiwara^2$ $Satoshi Sekine^2$

(1) Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara, Japan
(2) Rakuten Institute of Technology, New York, 215 Park Avenue South, New York, NY

yuta-h@is.naist.jp, {masato.hagiwara,
satoshi.b.sekine}@mail.rakuten.com

ABSTRACT

Most of the past error correction systems for ESL learners focus on local, lexical errors in a post-processing manner. However, learners with low English proficiency have difficulties even constructing basic sentence structure, and many grammatical errors can be prevented by presenting grammatical phrases or patterns while they are writing. To achieve this, we propose an integrated writing environment for ESL learners, called *phloat* to help such users look up dictionaries to find semantically appropriate and grammatical phrases in real-time. Using the system, users can look up phrases by either English or their native language (L1), without being aware of their input method. It subsequently suggests candidates to fill the slots of the phrases. Also, we cluster suggested phrases with semantic groups to help users find appropriate phrases. We conduct subject tests using the proposed system, and have found the system is useful to find the right phrases and expressions.

KEYWORDS: Writing Environment, Input Method, English as a Second Language, Suggestion of Patterns, Clustering of Candidates, Predicate Argument Structure.

1 Introduction

In the increasingly globalized world, opportunities to communicate with people who speak non-native languages are also increasing. English has established its position as the *de facto* lingua franca for many fields such as business and academia. This makes a large number of people worldwide to communicate in English as English-as-a-second-language (ESL).

Writing in English poses special challenges for those people. The major difficulties for ESL learners include lexicon, grammar or phrases, as well as articles and prepositions. A large number of systems and methods have been proposed to aid ESL learners for different aspects. For example, ESL Assistant and Criterion (Chodorow et al., 2010) is choosing and/or correcting appropriate articles and prepositions. Systems to assist learners in choosing the right verbs are also proposed, e.g. (Liu et al., 2011).

However, most of the systems are helpful only for local lexical or grammatical errors, and we believe a phrasal suggestion is needed to improve the English proficiency of ESL learners. Indeed, through a study on Japanese ESL learners' compositions, we found a significant number of errors which could be avoided if the learner knows appropriate phrases. Two examples from the study are shown in Example (1) (with some modification for simplicity):

> * The expense burden becomes your department.
> * It's usually the delivery of goods four business days. (1)

The first sentence, whose original intention is "Your department is responsible for the expense," is strongly affected by the first language (L1) expression なります *narimasu*, whose literal translation is "become." The author of this sentence translates the Japanese phrase into English literally, resulting in an almost incomprehensible sentence. The second sentence, with the original intention "It usually takes four business days to deliver the goods," does not follow any basic English structures. The author is assumed to have completely failed to put the words in the right order, even though he/she was successful in choosing the right words "delivery of goods" and "four business days."

These types of errors are extremely difficult to detect or correct by conventional ESL error correction systems. It is because they involve global structures of the sentence, and malformed sentences hinder robust analysis of sentence structures. Instead of attempting to correct sentences after they are written, we focus on *preventing* these types of errors even before they are actually made. As for the first erroneous sentence shown above, the error could have been prevented if we could somehow present a basic pattern "X is responsible for Y" to the author and let the author fill in the slots X and Y with appropriate words and phrases. Similarly, presenting phrases such as "it takes X to do Y" to the author of the second one could have been enough to prevent him/her from making such an error. In order to achieve this, there is a clear need for developing a mechanism to suggest and show such phrases while ESL learners are composing sentences.
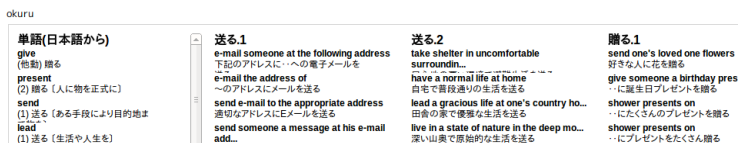

Figure 1: The response of input "okuru" (send)

We propose an integrated writing tool for ESL learners called *phloat* (PHrase LOokup Assistant Tool) to achieve this. The system is integrated within a text editor, and suggests English words and phrases based on the user input as shown in Figure 1. It also allows L1-based query input, i.e., if the user types a Romanized Japanese word, it shows all the words and phrases whose Japanese translations contain the query, as implemented also in PENS (Liu et al., 2000) and FLOW (Chen et al., 2012). In addition, the suggested phrases are classified and labeled based on their semantic roles (case frames). Figure 1 shows that, for an L1 input "okuru," which has several senses including "send something" and "spend (a life)," the system shows clusters of phrases corresponding to these senses. After the user chooses one of the phrases which contain slots to be filled, the system automatically shows suggestions for words which are likely to fit in the slots based on the context. This phrase clustering and slot suggestion help the user compose structurally more accurate, thus understandable English sentences.

We evaluate the effectiveness of our system based on the user experiments, where the ESL learners are asked to compose English sentences with and without the *phloat* system. The results were evaluated based on how grammatically accurate their English is (fluency), and how much of the original intention is conveyed (adequacy). Also the composition speed is measured. We will show that our system can help user find accurate phrases.

This paper is organized as follows: in Section 2, we summarize related work on English spelling/grammar correction and writing support systems. Section 3 describes the target, basic principle, and the design of our system. In Section 4, we describe the details of the implementation and the data we used. We elaborate on the experiment details in Section 5, and we further discuss the future work in Section 6.

2 Related Work

2.1 Post-edit Assistance Systems

There are a large number of post-edit systems for English spelling and grammar checking. ESL assistant, proposed by (Leacock et al., 2009) Web-based English writing assistance tool focused on errors which ESL learners are likely to make. It also shows parallel Web search results of the original and suggested expressions to provide the user with real-world examples. Criterion, developed by (Burstein et al., 2004) is another Web-based learning tool which uses an automatic scoring engine to rate the input learner's composition. It also shows detailed stylistic and grammatical feedback to the learner for educational purposes. Other than these, one can also find many other free or commercial English writing assistance systems including Grammarly[1], WhiteSmoke[2], and Ginger[3] to name a few. However, all these systems assume rather static input, i.e., focus on post-processing learners' compositions already finished. However, as stated in the previous section, many errors could be avoided by presenting appropriate feedback *while* the user is composing sentences.

2.2 Real-time Assistance Systems

Therefore, real-time assistance can be a more attractive solution for ESL error detection and correction. Recent versions of Microsoft Word[4] have a functionality to automatically detect spelling

---

[1]http://www.grammarly.com
[2]http://www.whitesmoke.com/
[3]http://www.getginger.jp/
[4]http://office.microsoft.com/en-us/word/

and elementary grammatical errors as the user types. AI-type[5] is an English input assistance software which helps users type at a higher rate by allowing partial matching of words. It also shows context-sensitive suggestions based on word n-grams.

PENS (Liu et al., 2000) is a machine-aided English writing system for Chinese users. Particularly noteworthy about the system is that it allows L1 (first language) input, that is, the system shows English translations for the user input in a Pinyin (Romanized Chinese) form. FLOW (Chen et al., 2012) is an English writing assistant system for Chinese, also allowing L1 input. Unlike PENS, FLOW further suggests paraphrases based on statistical machine translation to help users refine their composition. It is also helpful for writing in controlled natural languages to show real-time grammatical suggestions.

英文名文メイキング "Eibun Meibun Meikingu" (lit. meaning *Making excellent English*) (Doi et al., 1998) proposes an IME-type writing assistant for English. The system interacts with subsystems such as Japanese-English dictionary look-up, example sentence search, and *Eisaku Pen*, which converts Japanese input directly into English expressions. Google Pinyin IME[6] has support features for Chinese ESL learners, including integrated dictionary look-up, L1-based input support, and synonym suggestion. The same kind of L1-based dictionary look-up is also integrated in many IMEs, such as ATOK[7] and Google Japanese IME[8] for Japanese, Sogou[9], and Baidu Input Method[10] for Chinese. Some of those systems also support fuzzy matching with erroneous input, and suggestion of frequent phrases.

In a somewhat different line of research, controlled natural languages also benefit from writing support tools. AceWiki (Kuhn and Schwitter, 2008), which is a semantic wiki making use of a controlled natural language ACE, also provides an interactive writing support tool which automatically suggests subsequent word candidates as the user types.

Our proposal falls in this category. Compared to the previous systems, our tool is focus on phrase suggestion on top of the useful features developed in the past.

2.3 Phrase Search Systems

Phrase search plays an important role in English translation and composition. Several projects have been conducted for storing and searching useful English patterns such as "there is a tendency for [noun] to [verb]" (Takamatsu et al., 2012; Kato et al., 2008; Wible and Tsao, 2010). However, most of the phrase search systems require rather high level of English proficiency to use mainly targeted at technical writing. Also, they are not integrated in a text editor, or do not allow L1-based input, leaving a significant room for improvement when used in practice.

2.4 Translation Support System

ESL writing assistant is closely related to translation support systems because human translators often have to refer to a wide range of resources such as dictionaries and example sentences. To name a few of a wide variety of translation support systems, TransType2(Esteban et al., 2004) and TransAhead (Huang et al., 2012) suggest candidate words and phrases in the target language based

---

[5]http://aitype.com
[6]http://www.google.com/intl/zh-CN/ime/english/features.html
[7]http://www.atok.com/
[8]http://www.google.co.jp/ime/
[9]http://pinyin.sogou.com/
[10]http://shurufa.baidu.com/

on automatic translation of the source sentences. TWP (Translation Word Processor) (Muraki et al., 1994; Yamabana et al., 1997) is another translation tool which support composes the target sentence in an incremental and interactive manner.

3 Overview

3.1 Motivation

The proposed system aims to help users who are not necessarily good at English, especially in a writing form. For them, writing English is not an easy task; looking up dictionary, finding the right phrase, taking care of grammars and so on. It takes a lot of time for them to compose sentences. In such a situation, real-time assistance can be a promising help for them, because it saves a lot of their time spent on dictionary look-up. Additionally, because they may construct sentences based on L1 influences, the real-time assistance can prevent mistakes which may otherwise be difficult to correct by guessing what the users intended to say in the first place.

As we mentioned in Section 2, several systems have been proposed to address this issue. AI-type and FLOW suggest subsequent phrases based on n-gram statistics and machine translation, respectively. PENS and Google Pinyin IME suggest words corresponding to the L1 input.

However, these systems have two problems. First, they are not aiming at the users whose English proficiency is really low. As shown in Example (1), simple concatenation of partial translations does not necessarily produce grammatical sentences which express what they originally intended[11]. Second, although previous systems simply show candidates in a single column, it is difficult to find appropriate phrases when the number of candidates is really large.

In order to solve these problems, we propose an integrated writing environment for ESL learners called *phloat* (PHrase LOokup Assistant Tool).

3.2 System Overview

The proposed system works on a simple editor. Authors can write English sentences, as he/she does on a regular editor. Also, on top of English input, the author can type Romanized Japanese words when he/she does not know what to write in English. The system searches corresponding words in both languages, and displays the information in real-time.

For example, Figure 1 in Section 1 shows how the system supports when the user types "okuru" (which means "send" or "spend" in English). On its left, it displays the word translation candidates, and on its right (three columns in the figure), phrasal suggestions for "okuru" in three clusters are shown with Japanese translations. In this manner, the author can choose the appropriate word or phrase which matches the intent of the author.

If the author's intent is to write "sending email", the author can click the corresponding phrase (in the example, the second phrase of the first cluster of "okuru"). This action replaces the user input "okuru" with the corresponding English phrase "email the address of". As we know that we need to fill the slot of "address of", the system suggests the possible fillers of this slot (Figure 2).

The system works even the input is partial Japanese (Figure 3) or a part of English phrase (Figure 4). It also shows the suggestion for a combination of two Japanese words (Figure 5).

In the next subsection, we will summarize the notable features of the system.

---

[11]AceWiki indeed supports and ensures grammatical composition, although semantic consistency is not guaranteed.

e-mail the address of |

文脈から(英語)
their respective
the most
the world

Figure 2: The suggestion for the slot in "okuru"

okur

単語(日本語から)
over
(副) 送られて
give
(他動) 贈る
given
(形) 贈られた

Figure 3: The response to input "okur" (A prefix meaning "send" in Japanese)

get forg

単語(英語から)
forget
(自動) 思い出さないようにする
forgot
(動) forgetの過去・過去分詞形
forgotten
(動) forgetの過去分詞形

フレーズ(英語から)
get forgotten in the process
その過程で忘れ去られる
get forgetful
忘れっぽくなる
get forgotten
忘れ去られる
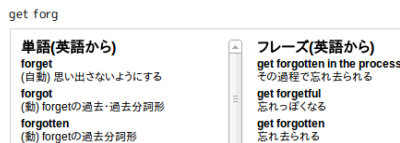
Figure 4: The response to input "get forg"

nimotsu okuru

単語(日本語から)
give
(他動) 贈る
present
(2) 贈る〔人に物を正式に〕
send
(1) 送る〔ある手段により目的地ま

フレーズ(日本語から)
stand in a long line to mail a pack
郵便局で荷物を送るために長い列に
send a package by express mail
荷物を速達便で送る
send a package via airmail to
～に航空便で荷物を送る
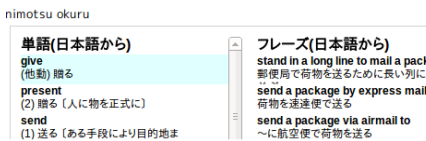
Figure 5: The response to input "nimotsu" (package) and " okuru" (send)

3.3 Features

Phrase Suggestion It suggests English phrases, on top of English words. As we have described, phrases could be better to be used in order to avoid unrepairable mistakes. Also, as the phrases are accompanied by Japanese sentences, the author may find the right phrase if it exists in the candidates. The phrases are semi-completed, natural phrases for native speakers like the ones shown in Table 2.

Slot Suggestion After the user chooses one of the candidate phrases, the system subsequently suggests candidates to fill in the slots of the phrase (Figure 2). These slot candidates are generated from the context of the suggested English phrase and its Japanese translation. This enables users to complete a long expression just by choosing a phrase and filling out the blanks in it.

Semantic Cluster Suggestion Since the system allows L1-based input in Romanized scripts, it results in a large number of phrase candidates which (probably partially) contain the L1 query, many of which belong to different semantic clusters. For example, Japanese verb "okuru" has at least two senses, "to send somebody something", and "to spend a life". Because our phrase list does not include sense knowledge, we need to group the phrases based on senses for the authors to find the appropriate one easily. The system suggests candidates with semantic groups (Figure 1), arranged it in multiple columns.

Flexible Input Words and phrases are suggested to the authors without being aware of their input methods; English or Romanized Japanese. Otherwise, the authors would have to switch their input method between the English direct input mode and the Japanese Kana-Kanji conversion mode, which is very laborious. The inputs in the Romanized Japanese are converted back to Kana phonetic characters (Figure 1) to find the candidates.

In addition, we implemented incremental search using the word prefix (Figure 3), making it unnecessary to type the complete words. This is the same for English input (Figure 4).

Search by Two Japanese Words In some cases, users would like to narrow down the phrases using two keywords, typically verbs and their arguments. For example, one can type "nimotsu okuru" (lit. package send) to narrow down phrases which are related to sending some packages, as illustrated in Figure 5.
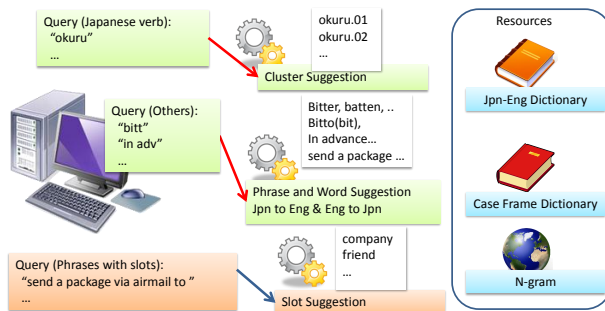
4 Implementation

Figure 6: Overview of the System

4.1 System Components

We illustrate the overview of the system in Figure 6. The system is comprised of three components - Normal word/phrase look-up, cluster suggestion, and slot suggestion. Since the system is unable to know the type of query (Romanized Japanese or English) from the input character sequence in advance, the system executes all the following search procedures every time the user types a character. The system receives 30 characters surrounding the carets as a query, which is in turn used for dictionary lookup:

- Looking up by a prefix of a Japanese word in Romaji
  (e.g.) *hashi* → chopsticks (hashi), edge (hashi), post (hashira)...
- Looking up by a prefix of a Japanese phrase in Romaji
  (e.g.) *nimotsu* → carry an armload of packages (nimotsu wo yama no youni kakaete iru)...
- Looking up by a prefix of an English word
  (e.g.) *cong* → congregation, congenital, congressman...
- Looking up a by a prefix of an English phrase
  (e.g.) *in adv* → in advance of ∽, in advanced disease...

After the look-up, all the results from each of the above search are shown. When the input is a Japanese verb, the phrases returned by the dictionary look-up are shown in semantic clusters. After the user chooses one of the phrases, the slot suggestion component is invoked to show suggestions.

All the components consult a Japanese-English dictionary. To achieve efficient look-up, all the entries in the dictionary are indexed with all the possible prefixes of their English translation and all the possible prefixes of tokens (except for particles) in Romaji of their sense in Japanese. For example, a phrase "carry an armload of packages" is indexed as shown in Table 1 [12].

	original string	prefixes
Japanese	nimotsu wo yama no youni kakaete iru	n, ni, nim, nimo, . . . , y, ya, yam, yama, . . .
English	carry an armload of packages	c, ca, car, carr, carry, carry a, carry an, . . .

Table 1: The Index of "carry an armload of packages"

---

[12] The variations, such as "ふ" (fu, hu) and "し"(shi, si) in Romanized forms of both Japanese senses and users input are all normalized by exploiting the Romaji table by Microsoft (http://support.microsoft.com/kb/883232/ja)

To avoid slow response and annoying users with too frequent suggestion, the first two searches (Japanese word/phrase look-up) are invoked only when the query is two or more characters long, and the last two are invoked only when the query is four or more characters long.

The system shows the set of lists, each of which is obtained from each component. The candidates in each list are sorted in a descending order of the language model score. That is, we simply ranked words depending on their unigram frequencies, and phrases on the language model score divided by the number of tokens in the phrase. We adopted stupid backoff (Brants et al., 2007).

To make it easier for users to choose a candidate even among a long list of too many candidates, the system groups them up by their meanings as described below.

Semantic Clustering Although it is ideal for the system to be able to cluster any phrases which match the user query, we simply performed clustering only for verbs because it is relatively easy to define what the "sense" of verb phrases are, given an appropriate lexical knowledge. Note that we could perform similar clustering on nouns as well using resources such as (Sasano and Kurohashi, 2009).

Clustering is performed only when the user inputs a Japanese verb such as "okuru," "nageru" using a case frame dictionary, which looks like the one illustrated in Table 3. The dictionary contains what kind of arguments each predicate takes, and the frequency is assigned to each argument meaning how often the argument is used in the case frame (predicate sense).

To each phrase including a single verb, we assigned a case frame[13] in the following way. First, we analyze the predicate argument structure of the phrase and obtain its arguments, by using a Japanese dependency parser, as stated below. We then sum up the frequencies of all the argument by consulting the case frame dictionary, if the analyzed case frame has any arguments. Finally we select the case frame with the highest frequency and assign it to the phrase.

For instance, the Japanese verb "送る" has two case frames, each of which corresponds to "to send" and "to spend", respectively, in the dictionary, and suppose we want to determine which case frame, 送る.1 or 送る.2. a phrase "メールを人に送る" (send an e-mail to a person) belongs to.

In the phrase, the predicate has two arguments, "メール" (mail) with accusative case and "人" (person) with dative case. The case frame 送る.1 has both of the arguments as its components, and its score is computed as the sum of their frequencies $(168 + 107022)$. On the other hand, the phrase only has one argument for the case frame 送る.2, which is "人," and its score will be the argument's frequency (80). Finally, because 送る.1 has a higher score than 送る.2, we regard the phrase belongs to the cluster 送る.1.

The system also suggests candidate to fill in the slot in phrases after the user selects a phrase in the suggested candidates. We describe how to obtain this below.

Slot Suggestion The system suggests candidates to fill in the slots in two ways.

First, it uses N-gram statistics to obtain commonly used phrases around the slots. Take a phrase "send a package via airmail to ～" (～に航空便で荷物を送る) for instance, which has "via airmail to" as the left context of the slot "～". We can obtain most common phrases which follow the words, e.g., "company," "friend," etc. by looking up the statistics.

We also use the right-hand-side context when available. The context length is limited by N, where N is the maximum length of N-grams contained in the database. Suppose that $N = 5$ and

---

[13]We discarded other phrases (e.g., phrases with two or more verbs) for clustering for simplicity

context $\cdots w_{-3}w_{-2}w_{-1} \sim w_1 w_2 w_3 \cdots$ is given, we can look up the N-gram database by multiple queries like $w_{-3}w_{-2}w_{-1}**$, $w_{-2}w_{-1}**w_1$, and so on[14]. We merged the multiple sets of phrases obtained in this way, and sorted them in the order of their frequencies.

Second, it uses a case frame dictionary to obtain plausible nouns which are likely to be filled in the slots. Taking the same phrase "send a package via airmail to $\sim$" for example, the case frame can be assigned in the same way as we described previously using the Japanese translation, and we know that the slot "$\sim$" has the accusative case (に *ni* case). Now we can look up the case frame dictionary and obtain the candidates using the case frame and the case, showing them in the order of frequency.

4.2 Data and pre-processing

Eijiro We used words and phrases contained in the Japanese-English dictionary Eijiro version 134 (released on May 23th 2012) [15]. This is a very large database of English-Japanese translations developed by the Electronic Dictionary Project. It can also be looked up at "Eijiro on the web"[16]. It is one of the most popular English dictionaries in Japan and it is accessed by over two million people in a month and searched over a billion times in a year[17]. It includes over 330,000 words, 1,434,000 phrases which contain no slots, and 256,000 phrases which contain one or more slots.

We automatically annotated the Japanese translations of phrases with part-of-speech tags using MeCab 0.994 [18] with IPA dictionary 2.7.0-2007080 [19] and parsed with dependency structures using CaboCha 0.64 [20].

In order to make the inversed index, we converted Japanese translations into Romaji, and obtained predicate argument structures. We regarded words which depend on a verb as arguments of the verb. Some samples of the predicate argument structure analysis are shown in Table 2.

Patterns	Predicate argument structures
〜に関する質問を（人）に E メールで送る	Verb: 送る Accusative:人
e-mail someone with one's questions regarding 〜	By:E メール
〜のアドレスにメールを送る	Verb:送る
e-mail the address of 〜	Accusative: メール Dative:アドレス
〜を E メールで送る	Verb:送る
send 〜 by e-mail	Accusative:〜 By:E メール

Table 2: Samples of Eijiro and their results of predicate argument structures analysis

Kyoto University's Case Frame We used Kyoto University's case frame data (KCF) ver 1.0 (Kawahara and Kurohashi, 2006)[21] as a Japanese case frame dictionary for slot suggestions and clustering. KCF is automatically constructed from 1.6 billion Japanese sentences on the Web. Each case frame is represented by a predicate and a set of its case filler words. It has about 40,000 predicates and 13 case frames on average for each predicate. We show some entries in KCF at Table 3.

Web 1T 5-gram We used Web 1T 5-gram Version 1 [22] which includes unigrams to five-grams collected from over 95 billion sentences on the Web for two purposes, for slot suggestion and

---

[14]Here $*$ denotes a wildcard.
[15]http://www.eijiro.jp/
[16]http://www.alc.co.jp/
[17]http://eowp.blogspot.jp/2011/12/on-web2011.html
[18]https://code.google.com/p/mecab/
[19]http://sourceforge.jp/projects/ipadic/
[20]https://code.google.com/p/cabocha/
[21]http://www.gsk.or.jp/catalog/GSK2008-B/catalog_e.html
[22]http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2006T13

送る.1	nominative	人:168, 私:152, ファン:89, 観客:80, 皆:73, < 数量 > 人:71, 誰:70, ...
	accusative	メール:107022, メッセージ:34957, エール:14356, 情報:14048, 写 9047, ...
	dative	方:3024, 友達:2679, アドレス:2492, 携帯:1704, 人:1557, 客:1443, < 数量 >9,
送る.2	nominative	人:80, 人々:62, 子供:55, 生徒:55, 学生:41, 自分:35, 方:29...
	accusative	生活:101316, 人生:19631, 余生:2001, 暮らし:1563, 寮生+活:57 ...
	dative	実際:39, < 補文 >:38, 基:19, 風:16, 通り:12, よう:11, 中:10, 普通:10, 中心:8...

Table 3: Entries of Kyoto University's case frame data

candidate ranking, as discussed previously. For slot suggestion, we eliminates candidates which include symbols such as ".", "?", and "$< /S >$". We indexed this with Search System for Giga-scale N-gram Corpus (SSGNC) 0.4.6 [23] to achieve fast look-up.

5 Evaluation

5.1 Methods

We conducted user experiments to evaluate the effectiveness of the proposed system. The subjects for the user test are 10 Japanese ESL learners whose English proficiency is intermediate level. Their English proficiency is measured by TOEIC [24] score, and their average TOEIC score was 754. We showed them two sets of English composition problems consisting of e-mail response writing, free picture description, and Japanese-to-English translation, and asked the subjects to write answer English sentences with and without using the proposed system. The problem sample is shown below (the actual problems were presented in Japanese):

We chose the two pictures shown in Problem 2 from the ones at Flickr [25] with a CC (Creative Commons) License, making sure that the pictures somehow involve states/actions of animals[26] and humans[27]. The sentences in Problem 3 were chosen randomly from Tatoeba project[28], excluding too simple or too complex sentences.

Instruction

- Please answer the following problems and write English sentences USING THE SYSTEM. (For the system group)
- Please answer the following problems and write English sentences WITHOUT using the system. You can use your favorite editors or word processor software, unless they have spell check functionality. You can freely consult any Japanese-English / English-Japanese dictionaries, such as "Eijiro on the web"[29]. (For the baseline group)

Problem 1: E-mail composition You placed an order for a bag from an oversea shopping website. But you found out that the item was partially broken and had some stains on it, and would like to exchange it for a new one or to return it. Fill in the blank below and complete the e-mail. Your composition will be evaluated in terms of: 1) how accurate your choice of English words and grammar is, and 2) your intension is conveyed to the recipient in full.

---

[23] http://code.google.com/p/ssgnc/
[24] https://www.ets.org/toeic
[25] http://www.flickr.com/
[26] http://www.flickr.com/photos/tjflex/233574885/
[27] http://www.flickr.com/photos/yourdon/3386629036/
[28] http://tatoeba.org/
[29] http://www.alc.co.jp/

Problem 2: Picture description Please describe the following picture using five or less English sentences. Your composition will be evaluated in terms of: 1) how accurate your choice of English words and grammar is, and 2) how accurately the reader of your description can reconstruct the original image.

Problem 3: Japanese-English translation Please translate the Japanese sentences in Table 4 into English. Your translation will be evaluated in terms of 1) how accurate your choice of English words and grammar is, and 2) how much of the original Japanese meaning is preserved.

Japanese	A sample of translation
彼はローマに行ってたくさんの古い建物を見た。	He went to Rome, where he saw a lot of old buildings.
彼女は約束を破ったといって彼を責めた。	She accused him of having broken his word.
この本はかつてはベストセラーだったが、今は絶版になっている。	This book, which was once a best seller, is now out of print.
紫のじゅうたんはこの赤いカーテンと調和しないだろう。	A purple carpet will not go with this red curtain.
誰かが間違って私のかさを持っていったに違いない。	Someone must have taken my umbrella by mistake.

Table 4: Japanese sentences of Problem 3

5.2 Scoring

We prepared two identically structured problem sets with different contents, and also divided the subjects into two groups with their TOEIC average scores as closer as possible. We asked the former group to solve Problem Set 1 with the system and Problem Set 2 without the system. This order is inverted for the latter group, i.e., they solve Problem Set 1 without the system and Problem Set 2 with the system, to cancel out the learning effect. We also measured the time taken to complete each problem set.

After completion of the test, we asked two native speakers of English to grade the subjects' composition. The grading was based on two measures, fluency and adequacy, which is done for each problem (each sentence for Problem 3), based on the following rubric:

Fluency How grammatically accurate the sentence is as English (in terms of grammar, word choice, etc.) (This only looks at grammaticality. For example, writing "how are you?" to describe a picture does not make any sense, but it is completely grammatical, so Fluency will be 5.)

Rating	Description
5	fluent (native-speakers can write this kind of sentences, possibly with some unnaturalness or awkwardness)
4	completely acceptable (the meaning is understandable with few errors; non-nativeness can be suspected)
3	acceptable (the meaning is understandable, the structure follows basic English grammar, with some non-critical errors e.g., spelling errors, articles, prepositions, word choices etc.)
2	unacceptable (have some (possibly partial) serious errors which may make comprehension difficult e.g., non-existent words, basic word order, etc.)
1	completely unacceptable (the form of the sentence has serious flaws which may make comprehension impossible)

Table 5: Criterion of rating fluency

Adequacy How much of the original intention is conveyed. (This only looks information. For example, writing "on floor cat" is not grammatical at all, but it might be enough to convert the meaning "there's a cat on the floor." In this case, some insignificant information is dropped, like number and tense, so adequacy will be 4)

The final fluency/adequacy score is calculated as the weighted average of each score, with Problem 1 being weighted by a coefficient of 2.0 and Problem 2 by 5.0. Finally, we also asked the subjects for any feedback and comments on the system usability.

| Rating	Description
5	full information conveyed (80%-100% of the original information is conveyed.
	The originally intended sentence can be reconstructed by simple paraphrasing)
4	most information conveyed (60%-80%)
3	half of information conveyed (40%-60%, including opposite meaning, e.g., dropped 'not' etc.)
2	some information conveyed (20%-40%)
1	almost no information conveyed (0%-20%)

Table 6: Criterion of rating adequacy
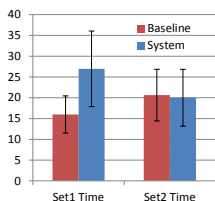
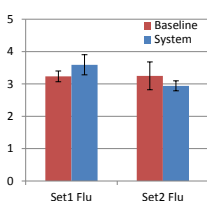Figure 7: Comparison of the averaged time (minute)
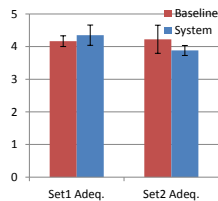
Figure 8: Comparison of the averaged fluency

Figure 9: Comparison of the averaged adequacy

5.3 Results

Figure 7 compares the time taken to finish each problem set when the system is used ("System") and not used ("Baseline"). The result is mixed, where it took a far more amount of time for Problem Set 1 when the system is used, while it shows little difference for Problem Set 2. In particular, we observed a few subjects find difficulties getting used to using the *phloat* system, doubling the time taken to complete the test. This shows that, although the system is designed to be as intuitive as possible, the familiarity with the system greatly affects one's writing efficiency, and this leaves us some room for improvements in terms of both user interface and word/phrase search relevancy.

Figure 8 and 9 compare the overall fluency and adequacy scores for both System and Baseline. Again, the result is mixed, where System is scoring higher for Problem Set 1 while Baseline is scoring higher for Set 2.

Some subjects claimed that it reduced their burden that they can easily look up some unfamiliar word such as ぶち *buchi* (spotted, tabby) using the system, which obviously reduced the time and/or helped increase the fluency.

On the other hand, after analyzing the result in detail, we found that there is one particular problem which scored badly on average for System, which is to translate the following sentence into English: "彼女は約束を破ったといって彼を責めた。" (She accused him of having broken his word.) The answers which the System subjects wrote for this problem include: " * She pilloried him to break faith with her." and " * She berate him to break faith with her." Both "pillory somebody for ..." and "berate someone for ..." come up as the search result for a query "semeru" (accuse), even though they are somewhat rare expressions to describe the original intention. Apparently, the users without enough English proficiency had no clues to decide which candidates are the most appropriate ones, resulting in non-optimal word choices. This phenomenon was also seen for a problem where subjects are required to translate "屈する" *kussuru* (be beaten, submit to). Some chose inappropriate word "bow" which came up as the first result. Notice that this problem can also happen when learners consult general dictionaries, and the fact that the *phloat* system facilitates the dictionary look-up made the issue even worse.

5.4 Advantages of the System

From the evaluation of the system by the ESL learners, we found several examples where the system was very helpful.

Change the time of meeting to X o'clock This phrase can be found in the phrase candidates when the user search "kaigi (meeting)" and "henkou (change)". This problem made one of the biggest differences between the results by the subjects who used the tool and the results by the subjects who did not use the tool. 4 out of 5 subjects who used the tool wrote this or similar phrases, whereas the subjects who did not use the tool wrote the followings: "change the meeting from 11am" (does not say "time of the meeting"), "alter the starting time to 11am", or expressed in very awkward structure although we can capture the meaning. It is evident that people had difficult time to construct the sentences without suggestions. This proves that if the system shows the phrases which exactly matches the intent, the tool is quite useful.

Wasteful expenses/spendings by the government This phrase and its variant are used by three subjects who used the tool. It is one of the phrase candidates when you search "shishutu (expense)" and "muda (waste)". This phrase also seems a difficult one to make by the subjects without the tool, maybe because it is a terminology in a special domain.

Comply with the advice We believe this phrase is not very easy for Japanese subjects to come up with, but the system suggests it when you search "shitagau (comply)" and "chukoku (advice)". Most subjects without the tool wrote "follow the advice", which is perfectly OK, but it is an interesting discovery for us that the tool could be useful to extend one's vocabulary by suggesting unfamiliar words.

6 Future Work

6.1 Flexible Match

In the current system, nothing is shown when the system cannot find any phrases which match input queries. It is mainly because the current system requires an exact match, not because the coverage of the Eijiro is limited. Even word variations, such as plural, past-tense and so on are not handled. It could increase the number of matched phrases by including word variance match, and we believe a proper ranking system is needed if it causes a problem. When there is no match, we may want to try fuzzy matching. It is very likely that the ESL learners may not remember the spelling correctly, the fuzzy matching can be helpful. In order to increase the number of the prepared phrases, adding sentences generated by machine translation may also be useful. However, it requires us to judge the accuracy of the candidates. Related research on interactive translation is conducted by (Kim et al., 2008; Tatsumi et al., 2012).

6.2 Fine Grained Selection

The subtle nuances of phrases are very difficult for ESL learners. Examples include "home" and "house," "at last" and "finally," and "must" and "have to." It may be difficult for the system to solve it by the context. One idea to solve the problem is that the system asks the author by showing some examples which use such phrases.

6.3 Smarter Suggestion using Contexts

The current system only looks at the keywords up to two words, and more context aware mechanism might be helpful to make smarter suggestions.

For example, a proper part-of-the-speech can be guessed based on the context. For queries after a verb such as "I stopped *okuru nimotsu (send package)*", noun phrases should be more likely than verb phrases. Or following "have" or auxiliary verb, past participle or basic forms of verbs are most suitable, respectively.

The choice of prepositions is very difficult task for ESL learners, because the choice is depending on the context (type of verbs and the role of fillers), and influenced by the L1 of the user. For example, the phrase "is covered" should be followed by "by" if the sentence is "the accident is covered by insurance", but by "with" if the sentence is "the mountain is covered with snow", even though the Japanese particles in both cases are the same.

Sometimes, we have to take into account collocations of phrases. For example, a Japanese word "ookii" can be translated into "large," "many" or "big," and the appropriate choice of the word must be done by considering the modified noun. For instance, large is the best modifier for population than theres. In order to suggest the right adjective, the system needs to consider the collocation.

6.4 Better Ranking Algorithm

The order of candidates in the suggestion list is very important, as the users,look at it from the top of the list. The current system gives ranks words and phrases based on the language model scores or frequencies of the candidates themselves (without sense or context into consideration).

For a Japanese word "Umi" (the main sense is "sea"), the dictionary lists "blue" as one of its senses (very minor usage in Japanese). However, because the frequency of "blue" in the English corpus is larger than that of "sea", the current system suggest "blue" at the highest rank. In order to avoid the problem, we need to have knowledge of major or minor senses for each word.

Conclusion

In this paper, we proposed an integrated writing tool for ESL learners, called *phloat* (PHrase LOokup Assistant Tool). It helps users who are not necessarily good at writing English with looking up words and phrases in a dictionary. Presenting appropriate phrases while the author is writing prevents from making serious mistakes which can't be fixed at post processing.

Our main contributions are the followings.

First, phrase suggestion is incorporated. These phrases can be searched by either English or Romanized Japanese, by one or more keywords. The users can easily find popular phrases, which are accompanied with the translation in their native language. In addition, it subsequently suggests candidates to fill the slots of the phrases. These suggestions enable users to complete a sentence just by choosing a phrase and fillers of the slots.

Second, we proposed clustering of suggested candidates with semantic groups. L1-based input sometimes results in a large number of phrase candidates. This helps users to find related phrases very easily.

Lastly, we evaluated the system asking subjects to write English sentences with or without the tool. It proved that the tool is quite useful when the system shows the phrases which exactly matches the intent and helpful to extend one's vocabulary by suggesting unfamiliar words.

References

Brants, T., Popat, A. C., Xu, P., Och, F. J., and Dean, J. (2007). Large Language Models in Machine Translation. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 858–867.

Burstein, J., Chodorow, M., and Leacock, C. (2004). Automated essay evaluation: the criterion online writing service. *AI Magazine*, 25(3):27–36.

Chen, M., Huang, S., Hsieh, H., Kao, T., and Chang, J. S. (2012). FLOW: A First-Language-Oriented Writing Assistant System. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics System Demonstrations*, pages 157–162.

Chodorow, M., Gamon, M., and Tetreault, J. (2010). The Utility of Article and Preposition Error Correction Systems for English Language Learners: Feedback and Assessment. *Language Testing*, 27(3):419–436.

Doi, S., Kamei, S.-i., and Yamabana, K. (1998). A text input front-end processor as an information access platform. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1*, pages 336–340.

Esteban, J., Lorenzo, J., Valderrábanos, A. S., and Lapalme, G. (2004). TransType2: an innovative computer-assisted translation system. In *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*, pages 94–97.

Huang, C.-c., Yang, P.-c., Chen, M.-h., Hsieh, H.-t., Kao, T.-h., and Chang, J. S. (2012). TransAhead: A Writing Assistant for CAT and CALL. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 16–19.

Kato, Y., Egawa, S., Matsubara, S., and Inagaki, Y. (2008). English Sentence Retrieval System Based on Dependency Structure and its Evaluation. In *Third International Conference on Digital Information Management*, pages 279–285.

Kawahara, D. and Kurohashi, S. (2006). Case Frame Compilation from the Web using High-Performance Computing. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*, pages 1344–1347.

Kim, C. H., Kwon, O.-W., and Kim, Y. K. (2008). What is Needed the Most in MT-Supported Paper Writing. In *Proceedings of the 22nd Pacific Asia Conference on Language, Information and Computation*, pages 418–427.

Kuhn, T. and Schwitter, R. (2008). Writing Support for Controlled Natural Languages. In *Proceedings of the Australasian Language Technology Association Workshop 2008*, pages 46–54.

Leacock, C., Gamon, M., and Brockett, C. (2009). User input and interactions on Microsoft Research ESL Assistant. In *Proceedings of the Fourth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 73–81.

Liu, T., Zhou, M., Gao, J., Xun, E., and Huang, C. (2000). PENS: A Machine-aided English Writing System for Chinese Users. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pages 529–536.

Liu, X., Han, B., and Zhou, M. (2011). Correcting Verb Selection Errors for ESL with the Perceptron. In *Computational Linguistics and Intelligent Text Processing*, pages 411–423.

Muraki, K., Akamine, S., Satoh, K., and Ando, S. (1994). TWP: How to Assist English Production on Japanese Word Processor. In *Proceedings of the 15th conference on Computational linguistics*, pages 847–852.

Sasano, R. and Kurohashi, S. (2009). A Probabilistic Model for Associative Anaphora Resolution. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 1455–1464.

Takamatsu, Y., Mizuno, J., Okazaki, N., and Inui, K. (2012). Construction of phrase search system for wiring english (in japanese). In *Proceedings of 18th Annual Meeting of The Association for Natural Language Processing*, pages 361–364.

Tatsumi, M., Hartley, A., Isahara, H., Kageura, K., Okamoto, T., and Shimizu, K. (2012). Building Translation Awareness in Occasional Authors: A User Case from Japan. In *Proceedings of the 16th Annual Conference of the European Association for Machine Translation*, pages 53–56.

Wible, D. and Tsao, N.-L. (2010). StringNet as a Computational Resource for Discovering and Investigating Linguistic Constructions. In *Proceedings of the NAACL HLT Workshop on Extracting and Using Constructions in Computational Linguistics*, pages 25–31.

Yamabana, K., Kamei, S.-i., Muraki, K., Doi, S., Tamura, S., and Satoh, K. (1997). A hybrid approach to interactive machine translation: integrating rule-based, corpus-based, and example-based method. In *Proceedings of the Fifteenth international joint conference on Artifical intelligence - Volume 2*, pages 977–982.