Evaluation of computerised reading-assistance systems for reading Japanese texts – from a linguistic point of view

Etsuko Toyoda
The University of Melbourne

For second-language learners to effectively and efficiently gather information from online texts in their target language, a well-designed computerised system to assist their reading is essential. While many articles and websites which introduce electronic second-language learning tools exist, evaluation of their functions in relation to the theories and empirical evidence revealed in research studies is still scarce. Due to the lack of information, second-language educators tend to leave learners to their own devices in choosing learning tools, which leads to ineffective and inefficient reading. The main aim of this article is to contribute to the development of a framework for evaluating computerised learning-assistance tools, by, as a case study, evaluating computerised reading-assistance systems for second-language readers of Japanese from a linguistic perspective. Firstly, criteria were defined through identifying crucial aspects of reading by reviewing reading-related literature. Two main criteria were information useful for reading comprehension (lexical coverage and accuracy of definitions) and information for the development of reading ability for future reading (explanations about content words, function words, kanji and components, structural and idiomatic information). The results of the evaluation revealed the strengths and weaknesses of different reading-assistance systems. This article then discusses implications for further research.

Introduction

In this globalised era, there is an ever-increasing need for an ability to gather information from the Web. For second-language (L2) learners to effectively and efficiently gather information from online texts in their target language, a well-designed computerised system to assist their reading is essential. This is particularly true when the L2 is orthographically, syntactically and semantically distinct from the first language (L1), an example of which is alphabetic-language-background learners reading in Japanese.

Being able to obtain information from the abundant materials on the Web is becoming increasingly crucial for knowledge construction. Distributed cognition theory emphasises that the construction of knowledge is realised through interactions between individual effort, other people, learning environment and tools (e.g., Ouyang & Stanley, 2014). From social constructivism perspectives, knowledge is seen as something that is created and shared in social settings (Gunawardena, Lowe, & Anderson, 1997). However, the construction of shared knowledge is often the result of the exchange of considered opinions based on researched information. The skills required for the comprehension of information in authentic texts in a foreign language have long been considered to be a product of years of studying vocabulary and grammar. However, many educators now think this to be only partly true. With fast-evolving technology, recent years have seen an abundance of electronic dictionaries, including computerised systems for reading assistance. Using a computerised reading-assistance system, learners of a foreign language should be able to gather information from online texts, the linguistic level of which exceeds the proficiency level of learners; they should not have to spend months or years memorising numerous linguistic items. This line of argument is also supported by Norman (1988), who argues that well-designed technological tools reduce the burden of learners needing to remember large amounts of information. However, an immediate problem foreign language learners and educators face is that they do not know which tool to use, because there are few evaluations of the usability of computerised learning-assistance systems and there are no established methodologies for such evaluation.

The aim of this article is to fill these gaps; it focuses on developing an evaluation framework for computerised learning-assistance tools that support L2 learning, by, as a case study, evaluating computerised reading-

assistance systems for L2 readers of Japanese. A study of dictionaries and learning assisting systems per se is as important as a study investigating learners who use the systems. While there are many articles and websites simply introducing language learning tools, evaluation based on linguistic theories and empirical evidence is scarce. Due to the lack of information, L2 educators tend to leave learners to their own devices in choosing learning tools, and how to use the chosen tool is left to the learners. Based on the results of a survey conducted on L2 learners of Japanese, however, Suzuki (2013) claims that students are in need of an orientation on types and features of systems. She reports that students are often not aware of the full functions of systems, or use multiple systems by trial and error, with a considerable cost in labour. Other studies indicate that, when provided with multifunctional e-dictionaries without guidance, L2 readers use almost exclusively the L1 definitions and ignore other information (e.g., Gettys, Imhof, & Kautz, 2001).

For a computerised reading-assistance system to be helpful for L2 readers, cosmetic features such as an attractive interface and easy navigation are not adequate. Crucially, it needs to provide L2 readers with necessary linguistic information such as accurate and complete definitions of words/phrases and grammar information for the purpose of low cognitive-load comprehension. In addition to the information for the current reading task, it should also provide information to build L2 readers' reading ability for future reading, thus reducing the reliance on reading-assistance systems. This article evaluates systems by focusing on the linguistic information.

This article will present an example methodology for evaluating computerised systems from a linguistics perspective. While acknowledging the importance of the evaluation of systems in terms of their interaction with users, the linguistic contents provided by the systems are critical. The following sections describe the methodologies and results of a study examining linguistic features (for both current and future reading) of existing computerised reading-assistance systems for L2 readers of Japanese using criteria developed based on theories and empirical evidence in reading-related research studies.

Empirical study

Reading theories and empirical evidence

In this section, in order to set criteria for evaluation from a linguistic point of view, I review literature to determine the type of information which should be provided in a computerised reading-assistance system. Reading processing can be divided into higher-level processing (i.e., comprehensions skills such as inference) and lower-level processing (i.e., decoding skills such as character/word recognition), and both levels of processing are essential for reading comprehension (e.g., Alderson, 2000; Koda, 1994). Recent reading models emphasise an interaction between a top-down processing (e.g., inferring and confirming the text and its components using the reader's schema – previous knowledge) and a bottom-up processing (i.e., processing from smaller units to larger units – words to sentences to paragraphs to the whole text) (Tateoka, 2005).

Lower-level processing

A number of researchers agree that automatic lower processing (i.e., processing of individual linguistic items) is critical (e.g., Koda, 1996). Effective and efficient processing of vocabulary and grammar information is essential for freeing mental resources for higher processing, such as interpretation (e.g., Grabe & Stoller, 2011). While the transfer of reading ability from L1 cannot be denied (Bernhardt, 2005), L2 linguistic knowledge (vocabulary and grammar) accounts for a larger amount of the variance in L2 reading performance, than L1 reading ability, at least until L2 readers reach a high level of proficiency (e.g., Jiang, 2011; Yamashita, 2002).

Some researchers claim that vocabulary instruction is far more important than grammar instruction (e.g., Horst, 2013; Saville-Troike, 1984). The size of vocabulary determines overall reading comprehension ability, particularly for low L2 proficiency readers (Horiba, 2012; Lee & Schaller, 1997). Laufer (1997) suggested the 95% lexical threshold in L2 reading. This means that for L2 readers to gain an adequate level of comprehension, they need to recognise 95% of the words used in the text that they read. Laufer adds that, for

pleasurable reading, 98% lexical coverage of texts is required. Hu and Nation (2000) claimed reading comprehension increased as the percentage of known words increased, based on the results of their study. A more recent study also revealed a relatively linear relationship between the percentage of vocabulary known and the degree of reading comprehension (Schmitt, Jiang, & Grabe, 2011). A computerised reading-assistance system should provide accurate semantic information on all words and phrases to enhance lexical coverage of texts. Although increasing lexical coverage from a very low level (e.g., the reader has few known words resulting in 10% coverage) to a sufficient level (e.g., 95%), using a reading-assistance system, should not be expected, a system should have accurate definitions and explanations of all the linguistic items ready for the system user. L2 readers with intermediate to advanced proficiency would probably benefit most from the use of a reading-assistance system. However, even beginner L2 readers trying to decipher an online TV guide, for instance, could also benefit from it.

Providing definitions of all linguistic items is not sufficient. Some researchers suggest that, without explicit instructions, L2 readers may grasp only partial vocabulary knowledge, such as one usage of the word's multiple usages (Hsu & Yang, 2013). Vocabulary studies suggest that breadth (i.e., the number of words) and depth (i.e., detailed knowledge about words) of vocabulary knowledge, and incidental and intentional vocabulary learning are all important for the improvement of such knowledge, and consequently reading skills (Nation, 2001). However, the sheer number of words makes it almost impossible for L2 readers to acquire the required breadth and depth of vocabulary knowledge within a limited time (Nation, 2006). To this end, a reading-assistance system should provide extra information to improve depth of vocabulary knowledge, in addition to fundamental information such as definitions of words for reading comprehension. However, currently, few e-learning systems provide opportunities for learning both breadth and depth of vocabulary knowledge (Hsu & Yang, 2013).

It is undeniable that, for a logographic language such as Japanese, lower-level processing is highly important (Ehrich, Zhang, Mu, & Ehrich, 2013). The processing of Japanese texts, having complex orthography and a large number of compound words including those comprising multiple kanji (Chinese characters used in Japanese), is difficult particularly for Japanese learners with no Chinese character knowledge. The Japanese vocabulary learning load has been shown to be much higher compared with other languages. Studies dealing with Indo-European languages suggest that vocabulary knowledge determines approximately 30% of one's reading ability (Bernhardt, 2005). However, in Japanese, the figure has been estimated to be higher than 40% (Komori, Mikuni, & Kondo, 2004), due to a large number of words written in kanji. In relation to this, the knowledge of sub-word units (kanji and their components) is also essential (e.g., Koda, 2002; Toyoda, 2007). According to the path model presented by Tamaoka and Yamada (2000), knowledge of functional components (i.e., components that carry useful semantic or phonetic information) is an important contributor to knowledge of kanji lexical orthography, phonology and semantics. A number of studies have found that there are strong correlations between reading comprehension and vocabulary knowledge, and between vocabulary knowledge and word recognition skills (e.g., Grabe, 2004; Horiba, 2012; Horst, 2013). In Japanese, having both breadth and depth of vocabulary knowledge and word recognition for words written in kanji is essential. Thus, L2 readers need to be provided with opportunities to learn about words and sub-word units (i.e., kanji and their components) while using a reading-assistance system.

Some researchers claim that grammar knowledge plays a pivotal role in reading comprehension (Jeon & Yamashita, 2014; Shiotsu & Weir, 2007). Since Japanese syntax is very different from that of Indo-European languages, knowledge of function words such as particles and auxiliaries can be crucial. In Japanese, particles are often omitted in casual writing (Lee, 2002; Nariyama, 2002). This ellipsis of particles can cause miscomprehension if L2 readers are not provided with some relevant knowledge for inference. Conjugations of verbs and adjectives, representing tenses, aspects and voices, can also be essential to the interpretation of the text. Myers (1994) argues that all grammatical information required should be entered into the dictionary alongside information of stems. The term *vocabulary* is generally used to refer to content words. However, it is apparent that providing the knowledge of both content and function words is essential.

Upper-level processing
Noun-modifying clauses can cause as much disturbance to L2 reading in Japanese as particle ellipses. Japanese noun-modifying phrases are often ambiguous because, unlike relative clauses in English or other similar languages, Japanese allows noun-modifying clauses to form without relative pronouns or verb conjugations, which typically give clues to relationships between modifiers and modified nouns (Matsumoto, 2007). Deciphering of noun-modifying clauses is often achieved through an aggregate of semantic and pragmatic factors (Matsumoto, 2007). Another potential miscomprehension is idioms and idiomatic phrases and expressions, which require historical or cultural knowledge (Pulido, 2007). Myers (1994) reports that advanced learners of English frequently encounter cultural-specific, or idiomatic, words which they cannot find in dictionaries. The situation is the same in Japanese. L2 readers of Japanese should be provided with some idiomatic information as well as Japan/Japanese specific information.

Evaluation of reading-assistance systems

Targets of evaluation
While many electronic dictionaries and translators exist, there are few reading-assistance systems for L2 Japanese readers. This is mainly due to the writing system of the Japanese language, which has no clear boundaries between words. This segmentation issue hinders a rich emergence of reading-assistance systems. For this study, the following four systems became the target of examination: Reading Chuta (hereafter referred to as Chuta; http://language.tiu.ac.jp/), Asunaro (https://hinoki-project.org/asunaro/), Rikai (http://www.rikai.com/perl/Home.pl), and WWWJDIC (http://nihongo.monash.edu/cgi-bin/wwwjdic?9T). These systems are for providing a learning environment where L2 learners can use Internet information for their Japanese study, and are specifically designed to enable L2 learners to read Japanese texts with hyper-glosses by simply pasting texts.

The main focus is their functions for English-speaking L2 readers of Japanese, although some of these systems are capable of assisting speakers of languages other than English. Other additional functions, such as displaying kanji information, are also taken into consideration for evaluation even if they are not part of the main reading-assistance function.

Evaluation materials
Three unmodified Japanese text segments containing all the information identified as essential for reading comprehension, namely, content words, function words, kanji compound words, structural information, and idiomatic information, in different styles of writing (plain, polite, and colloquial styles), were chosen as test passages. The three passages were pasted into the four reading-assistance systems, and the results of processed passages were compared.

Passage 1

> 「馬を水辺に連れていくことはできるが、水を飲ませることはできない」ということわざもあるように、やる気の低い部下の耳元で、いくらリーダーが「もっと全力で仕事に取り組め、手を抜くな」と声を張り上げたところで、そんなのは文字どおり、馬の耳に念仏だ。しかしながら、まともなリーダーに代わったら、俄然メンバーの眼の色が違ってきて、みんなやる気を出し始めたというようなこともよくある話だ。
> (Yoshikoshi, 2012)

Passage 2

面接のコツを知っていれば鬼に金棒です。就職活動において面接は、自分の人生に大きな影響を与える出来事です。面接のためのノウハウやコツを知ることが大切です。コツを知って、平常心をもって臨めば結果はついてくるでしょう。面接では話す内容だけでなく、どのように話すかが重要になってきます。聞かれた事に答えるだけでなく、進んで自己のＰＲや志望動機を話しましょう。面接において姿勢は重要な要素です。姿勢を正すことで、自分の自信が人事担当に伝わります。これは心理学の分野で実証されていることで、不思議と自分も自信を持つ効果があるのです。
(Shukatsu Navi, 2014）

Passage 3

なんせ、愛好しているわけだから。大枚をはたいて、多大な時間をかけているのだ。単に音が出りゃあいいってもんじゃないだろ。ただしこれ、きわめて主観的である。やたら大きくて周囲を圧倒するくらいのが好きな人。コンパクトで何かが凝縮した感じが好みの人。ツルツルな感触から離れられない人。ごっつい感じじゃないとダメな人。メカメカしいものにぐっと来るヤカラ、アールデコな様式美を持ったデザインにうっとりしちゃう御仁まで十人十色、蓼食う虫も好き好きである。
(Suzuki, 2013)

Evaluation criteria

The reading-assistance systems were evaluated for the following information, which has been identified, through reviewing reading-related literature, as crucial aspects of reading:

(1) information to assist L2 readers to comprehend the text (i.e., providing high-coverage and accurate itemised information)
(2) information to help L2 readers develop their reading ability for future use, namely, content word, function word, kanji and component, structural and idiomatic information.

Evaluation methods

For criterion 1, information to assist L2 readers to comprehend the text was evaluated in two steps:

Step 1 – Coverage was evaluated using numbers and percentages of highlighted characters (flagging that information is provided). As the definition of a *word* is ambiguous in Japanese (Kato, 2006), characters, rather than words, were counted in order to determine the proportion of coverage. Firstly, for each passage, total characters per passage were counted, after deleting all punctuation. Secondly, the highlighted characters (indicative of having a gloss) in each system were counted, and percentages were then calculated. For example, the sentence 馬を水辺に連れていく。has 10 characters in total, after removing the punctuation. The number of highlighted (underlined) characters is 7 (馬を水辺に連れていく); therefore, the coverage is 70% (7/10*100).

Step 2 – Accuracy of the given information was evaluated using numbers and percentages of *helpful for accurate comprehension* items. Firstly, the number of items was counted using the unit of segmentation in the system as the basis for counting. For example, while in Chuta and Asunaro, 連れて [accompany] and いく [go] are glossed as two separate items (as highlighted in 連れて いく), Rikai and WWWJDIC list them as one item, 連れていく [take someone along] (as highlighted in 連れていく). Thus, the number of items in 馬を水辺に連れていく is four (馬を水辺に連れていく) according to Chuta and Asunaro, and three (馬を水辺に連れていく) according to Rikai and WWWJDIC. The item 連れていく is a compound word comprising two verbs, 連れる and いく, and therefore, from the semantic perspective, could be divided into two. However, for the analysis, if a compound word was highlighted as one item, it was treated as such, because for the user of the system, it is one item to click to seek information. Secondly, those items with

glosses likely to assist reading comprehension were counted. In other words, those without definitions and those that would give misleading meanings (e.g., providing definitions of individual words instead of a definition of a set phrase as a whole) were not counted. The percentages of items with helpful glosses were then calculated. The determination of misleading glosses was conducted by two bilinguals, being the author and a research assistant. A third party was invited where there was a lack of consensus.

For criterion 2, whether or not a system offered the range of information useful for building depth of knowledge (the five aspects of information identified as important in literature) was investigated. Content words, function words and idiomatic expressions were found in the main reading-assistance pages of the systems. All the highlighted items provided with definitions, regardless of the *accuracy*, were classified into a content word, a function word, or an idiomatic expression, and respective numbers counted. The classification was conducted by the author and the research assistant. For example, in the sentence 馬を水辺に連れていく, there were three content word items (馬, 水辺, and 連れる) and one function word item (いく in ていく). In this study, adjunct words such as ていく, てみる, ておく were treated as function words, based on the definitions of word types given in Kato (2006), although this view may not be shared by some researchers. Another issue was how to treat 連れていく when it was highlighted and given a definition as one item, as in WWWJDIC. In this study, it was treated as one content word item, since it has its own definition as a stand-alone word. All in all, however, the classifications and the counting of items should be treated as indicative only. The search for kanji and structural information was extended beyond the main pages, as, except Rikai's kanji information, these were provided as separate functions.

Results

Information to assist L2 readers to comprehend the text

For all three passages, WWWJDIC had the highest cover rates (averages of 75.57%), though still far from 100% (see Table 1). The figures shown for Asunaro are the numbers of highlighted characters and the cover rates of such characters, not the numbers and rates of information provided. At present (as of April 2015) no dictionary information can be seen, although their documentation (Abekawa et al., 2004) suggests otherwise.

Table 1
Numbers of characters and percentages of information provided

		Total # of characters	Chuta	Asunaro	Rikai	WWWJDIC
Passage 1		173	122	113	71	136
	Coverage		70.52%	65.31%	41.04%	78.61%
Passage 2		245	152	147	120	186
	Coverage		62.04%	60.00%	48.98%	75..92%
Passage 3		205	143	121	80	148
	Coverage		69.76%	59.02%	39.02%	72.20%
Average			67.44%	61.44%	43.01%	75.57%

However, the coverage rates alone are not a good predictor of the usability of the reading-assistance systems. In the case of Asunaro, the highlighted items did not show any meanings or explanations. A few highlighted items in Chuta did not have glosses. In addition, the information provided can sometimes be misleading. If the used segmentation unit is not correct, the information (even if it is the correct meaning for that segmented item) can be confusing. For example, when an idiomatic expression is divided into smaller units, the information for each item may confuse the reader (馬 [horse], 耳 [ear], 念仏 [the chanted words] for 馬の耳に念仏 [not heeding what others say]). Several such cases were observed in Chuta and, to a lesser extent,

in Rikai. In WWWJDIC, over-segmentation of an idiomatic expression was rare, as it provided the definitions of idiomatic expressions as a whole. However, a few wrong segmentations of another sort were identified. For example, ということわざもある [there is a proverb saying …] should be segmented into という/ことわざ/も/ある. However, WWWJDIC mis-segmented it, and provided information for ということ, which is an amalgamation of という [saying] and the first half of ことわざ [proverb]. Table 2 shows the accuracy rates of Chuta, Asunaro, Rikai, and WWWJDIC for passages 1, 2, and 3. The accuracy of Asunaro is shown as 0% because none of the items was provided with a gloss, and therefore there is no way to judge its accuracy.

The results indicate that WWWJDIC is the most reliable of the four, with the largest coverage.

Table 2
Numbers and percentages of accurate items

Correct/Total # (types)		Chuta	Asunaro	Rikai	WWWJDIC
Passage 1		36/45	0/44	26/27	37/40
	Accuracy	80.00%	0%	96.30%	92.50%
Passage 2		43/54	0/59	40/42	47/48
	Accuracy	79.63%	0%	95.24%	97.92%
Passage 3		36/54	0/49	28/30	40/42
	Accuracy	66.67%	0%	93.33%	95.24%
Average		75.43%	0%	94.96%	95.22%

Information to help L2 readers develop their reading ability for future use
For the five required aspects of information (i.e., content word, function word, kanji, structural and idiomatic information), no system provides everything (see Table 3). The figures provided for content words, function words and idiomatic phrases are the number of types (rather than tokens) and are indicative only, as discussed in the previous section.

Chuta, and Rikai, in particular, provided information almost exclusively for the content words. Rikai provided kanji information as well as words. Asunaro showed nothing but examples (without translations). However, this system has a separate function of showing sentence structures in multiple ways (e.g., tree structure and box structure). WWWJDIC provided all but structural information.

Table 3
Five aspects of information identified as important in literature

	Information	Chuta	Asunaro	Rikai	WWWJDIC
Passage 1	Content word	YES 35 types	NO Examples only	YES 25 types	YES 29 types
	Function word	YES 3 types	NO Examples only	NO	YES 6 types
	Kanji	NO	NO	YES	YES Separate function
	Structural	NO	YES Separate function	NO	NO
	Idiomatic	NO	NO	YES 2 types	YES 5 types
Passage 2	Content word	YES 45 types	NO Examples only	YES 42 types	YES 34 types
	Function word	YES 3 types	NO Examples only	NO	YES 8 types
	Kanji	NO	NO	YES	YES Separate function
	Structural	NO	YES Separate function	NO	NO
	Idiomatic	NO	NO	NO	YES 6 types
Passage 3	Content word	YES 40 types	NO Examples only	YES 27 types	YES 32 types
	Function word	YES 3 types	NO Examples only	YES 1 types	YES 5 types
	Kanji	NO	NO	YES	YES Separate function
	Structural	NO	YES Separate function	NO	NO
	Idiomatic	YES 1 type	NO	YES 2 types	YES 5 types

Discussions on evaluation

A reading-assistance system should provide information to assist L2 readers to comprehend the text for the current use and to develop reading ability for future use. However, the evaluation of the above four systems revealed that none of these systems gives the L2 reader of Japanese sufficient information for both.

The coverage rates were all low. The highest information coverage (which WWWJDIC showed) is still far from 100%, and probably not adequate for full comprehension. Vocabulary studies repeatedly find a strong correlation between the percentage of words known in a text and the degree of reading comprehension (e.g., Hu & Nation, 2000; Schmitt et al., 2011). Higher coverage in a system means that words unknown to L2 readers are more likely to be covered, resulting in better comprehension.

Although the coverage rates may vary depending on a text, a reason for the overall low coverage in the reading-assistance systems seems to be that function words, such as particles and auxiliary words, and verb/adjective conjugations, are not adequately covered. As function words are context-dependent, generic explanation is often too abstract to be useful. Giving a description on how the function word is used in each context is impossible without taking surrounding words into consideration, for which we have to await natural language processing (NLP) researchers to succeed. Nevertheless, in order to understand an unmodified text, information concerning function words is crucial. Hori (2012) completed the compilation of all possible usages of function words and phrases taken from both spoken and written corpora, together with an example sentence per usage. The inclusion of such detailed information of functional items will improve the existing reading-assistance systems. However, explaining all usages of a function word with an example sentence appropriate for each usage may be problematic given space constraints (i.e., the screen space is limited).

None of the systems gives information about verb/adjective conjugations although WWWJDIC provides notes mentioning conjugation forms, such as "it is possibly the volitional form". This may be due to the fact that the Japanese conjugations are thought to be simple and systematic, and that there are already hundreds of websites showing basic conjugation patterns. However, understanding conjugations accurately is crucial for comprehension (Hazelbeck & Saito, 2012; Shiotsu & Weir, 2007). This is particularly true in the languages, including Japanese, in which meaning-determining information, such as potential, passive and causative forms, is expressed in the conjugation. Having conjugation information, including glossing in English, inside a system would assist L2 readers to read a text without being interrupted by the necessity of checking the form with other reference books. However, providing English meanings for conjugations is not simple in Japanese. Since Japanese is an agglutinative language, tense, aspect, voice and mood can all be expressed in seemingly one single conjugation. For example:

食べさせられていたようだ
tabesaserareteitayōda
[it appears (subject) was being forced to eat]

is an amalgamation of
食べ＋させ＋られ＋てい＋た＋よう＋だ
tabe + sase + rare + tei + ta + yō + da

In order to show the meaning of the whole phrase, the system needs to combine the meanings of all the segments.

The analysis of the three passages indicated that WWWJDIC had the highest accuracy. The strength of this system is that it provides information on multiword expressions (MWEs), such as idioms, fixed phrases, noun compounds, and compound verbs. Tanabe, Takahashi, and Shudo (2014) claim that MWEs, which occur in authentic sentences with unexpectedly high frequency, can cause serious problems in comprehension. It is estimated that half of the average English speaker's lexicon is MWEs (Jackendoff, 1997). Although an exact percentage in the Japanese lexicon is unknown, it is enormous (Tanabe et al. 2014). There are two ways of dealing with MWE issues. One is, as WWWJDIC does, to treat a whole MWE as a headword, and gloss it. The other way is, as McAlpine and Myles (2003) propose, to provide typical phraseology for a *node* word. Chuta, on its web dictionary page, presents a few common collocational patterns for some node words. However, Chuta falls short in providing information on function MWEs (e.g., かもしれない [might be], なければならない [must]) as well as single function words.

WWWJDIC contains a large number of compound words in its dictionary, thus providing the L2 reader with fairly adequate information to comprehend sentences. Chuta also glosses some compound words, although many are missing. A large volume of compound verbs (e.g., 見失う [look-lose], and 泣き止む [cry-stop]), which are a type of MWEs (Tanabe et al., 2014), is one of the features of the Japanese language. Major dictionaries typically include several thousand as entries, and the actual number being used is suspected to be

much higher due to their highly productive nature (Breen & Baldwin, 2009). The number of MWEs is rapidly increasing also due to the import of a large quantity of English words and the creation of new MWEs based on the loanwords (Breen, Baldwin, & Bond, 2012). The above three passages used for the comparison did not contain loan-MWEs. If they did, WWWJDIC would have performed well, as it has a large volume of such MWEs in its dictionary.

The fundamental problem that Chuta or other current reading-assistance systems (or anything that needs to decipher a free text) face is word segmentation. The identification of words in text usually needs to rely on morphological analysis for the two reasons: because there is no space to flag the start or end of a word in Japanese and because Japanese is an agglutinative language wherein a word is ambiguously defined. One advantage of the use of a morphological analyser is that all orthographic variants and all conjugational forms can be identified under the same morpheme (Den et al., 2007). However, most available Japanese morphological analysers use the smallest meaningful unit (i.e., a minimum grammatical unit that carries semantic distinction) for the segmentation of a text, often conflicting with the size of the unit, with which Japanese language learners are familiar (Yamauchi, 2008). There are several morphological analysers, and not all segment a sentence in the same manner. Even within a morphological analyser, the segmentation unit varies; mostly a strict morpheme, but sometimes larger. Although some morphological analysers offer a longer unit of segmentation, the unit is often not long enough for most MWEs. Thus, Chuta and Asunaro fail to provide information on most MWEs, although these systems seem not to be limited to the use of small units, but also medium units, to segment pasted text, given that some compound verbs are listed as headwords. To improve this situation, the development of large-unit morphological analysers is under way (e.g., Kozawa, Uchimoto, & Den, 2014).

On the other hand, WWWJDIC and Rikai use a greedy algorithm (searching for a longest match in the dictionary) and successfully find MWEs in text and display their respective information provided the MWEs are in their dictionaries. A downside of an inbuilt dictionary without feeding through a morphological analyser may be, as seen in WWWJDIC for passage 3, it sometimes fails to find the words/phrases in the dictionary if they appear in a text in a variant form, such as the use of kana instead of kanji (e.g., つくって instead of 作って), which could be resolved if an analyser were used. Rikai suffered from the same problem. Another type of error caused by a greedy algorithm is the above-mentioned mis-segmentation. Where と/いう/ことわざ /も/ある would have been the correct segmentation, WWWJDIC matched ということ as the longest match. In order to avoid this type of errors, Hazelbeck and Saito (2012) used a hybrid method for their glossing tool; converting the output of a morphological analyser to a format that is more natural for L2 readers of Japanese.

In relation to the existence of the above-mentioned five aspects of information identified as important in literature, Asunaro is the only system that provides structural information. The structural information is potentially useful, and in some cases, it helps L2 readers to see the relationships between the modifiers and modified nouns. NLP research studies have been conducted in this area using various types of Japanese dependency parsers (e.g., Kudo & Matsumoto, 2002). However, it seems it is still too early for the results of these studies to be implemented within a reading-assistance system.

Kanji information is provided in WWWJDIC and Rikai. In WWWJDIC, however, the relationships between words and kanji are insufficiently evident since the display of kanji information can only be done by selecting a function separate from glossing. Rikai, on the other hand shows kanji together with the words in which the kanji occurs, and therefore the relationships are more noticeable. With regards to the relationships between kanji and their components, WWWJDIC has a separate function to give some componential information of kanji. However, the semantic components provided in this system are treated as mere building blocks, rather than the components that may carry semantic information pertaining to the kanji. Besides the semantic components, a certain number of phonetic components (carrying useful phonetic information) have been identified (Toyoda, Arief, & Kano, 2013). Thus, the inclusion of useful components may assist L2 readers of Japanese to improve their reading ability. The results of a survey on dictionary use suggests that advanced L2

learners are aware of the importance of kanji and their components for reading and writing Japanese texts (Suzuki, 2013).

One desired function that none of the systems has is the display of multiple meanings of polysemes in the order of most to least appropriate for a given context. L2 readers often have difficulties choosing the right meaning out of many listings (McAlpine & Myles, 2003). Some suggestions include linking each polyseme with a possible collocational pattern, and then glossing it (McAlpine & Myles, 2003) and reminding L2 readers about meaning choices (which meaning they selected at the previous encounter) together with contextual information (context sentences from the previous encounter) (Dang, Chen, Dang, Li, & Nurkhamid, 2013).

Conclusion and implications

This study was not conducted to determine which system to use for L2 reading comprehension, nor to propose an ideal multifunctional reading-assistance system for L2 Japanese learners. The aim of this article was to contribute to the development of a framework for evaluating computerised learning-assistance tools. As a case study, this paper demonstrated an evaluation of computerised reading-assistance systems for L2 readers of Japanese based on linguistic theories and empirical evidence, which is still scarce.

Whether L2 reading would be greatly assisted, had we a perfect computerised reading-assistance system, is a completely separate issue that needs to be studied. A conceptual perfect system may not, of itself, enable L2 Japanese learners to read authentic texts independently. It is probable that L2 readers would be overwhelmed by too much information. Also, depending on the purpose of reading, some functions may be totally irrelevant or even a hindrance to the current task. Furthermore, there is probably a threshold proficiency level for L2 readers to be effectively assisted by such a multifunctional system. Even for capable L2 readers, training may be necessary for them to utilise the functions, particularly those provided for future reading. Where a system provides multiple reading-assistance functions, they need to be switchable (suppressed or otherwise) to suit the needs of the L2 readers. It is crucial that assistance be provided by L2 educators in the use of relevant functions.

Even if an ideal reading-assistance system is introduced with thorough guidance, it does not necessarily promise instant improvement in text comprehension in learners, as comprehension may be affected by various factors. For example, the four systems presented here have very different layouts and navigation. Factors such as these are of great importance in considering how L2 Japanese learners use these resources. Reading comprehension efficiency is also affected by the difficulty of the text, and by various human factors, including L2 proficiency, working memory capacity, cognitive styles, metacognitive skills, world knowledge, reading purposes, reading habits, reading strategies, personal traits, experience of using computerised tools, and motivation. These factors need to be considered when investigating how effective a reading-assistance system is for L2 reading comprehension.

The main aim of this article was to contribute to the development of an evaluation framework for computerised learning-assistance tools that support L2 learning, by, as a case study, evaluating computerised reading-assistance systems for L2 readers of Japanese. The scale of this empirical evaluation study is not broad enough to provide reliable evaluation results on the usability of the computerised reading-assistance systems, nor to claim that the methodology used here should be the methodology for evaluation. However, it demonstrates that evaluation criteria could be based on linguistic theories, and it also presents procedures for evaluation. This is significant because some computerised learning systems, such as emerging smartphone apps, which are supported by educational technology theories, are short on important linguistic information.

Detailed examinations of the systems through a L2 educators' point of view are also invaluable for system developers. Language learning tools are critical for L2 learning, and yet L2 language educators have long been relying on developers to design and create them. More examinations and evaluations by L2 educators will be valuable sources of advice for guiding improvements of future systems. In this respect, this article

paves the way for future research by demonstrating a methodology supported by research in linguistics/applied linguistics.

Acknowledgements

This study is part of a large project funded by Hakuho Foundation.

References

Abekawa, T., Yagi, Y. Totsugi, T., Fu, L. Bilac, S. Okumura, M. & Nishina, K. (2004). 読解支援システムのための言語非依存フレームワーク構築 [Development of language-independent framework for reading support system]. In *Proceedings of 10th Conference of Natural Language Processing* (pp. 144–147). Tokyo: Association of Natural Language Processing. Retrieved from http://lr-www.pi.titech.ac.jp/~abekawa/pdf/asunaro04a.pdf

Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Bernhardt, E. (2005). Progress and procrastination in second language reading. *Annual Review of Applied Linguistics, 25*, 133–150. doi:10.1017/S0267190505000073

Breen, J., & Baldwin, T. (2009). Corpus-based extraction of Japanese compound verbs. In L. Pizzato & R. Scwitter (Eds.), *Proceedings of the Australasian Language Technology Workshop* (pp. 35–43). Australasian Language Technology Association. Retrieved from http://www.alta.asn.au/events/alta2009/pdf/ALTA2009_05.pdf

Breen, J., Baldwin, T., & Bond, F. (2012). Segmentation and translation of Japanese multi-word loanwords. In P. Cook & S. Nowson (Eds.), *Proceedings of the Australasian Language Technology Workshop* (pp. 61−69). Australasian Language Technology Association. Retrieved from http://www.alta.asn.au/events/alta2012/proceedings/pdf/U12-1009v2.pdf

Dang, T. D., Chen, G.-D., Dang, G., Li, L.-Y., & Nurkhamid (2013). RoLo: A dictionary interface that minimizes extraneous cognitive load of lookup and supports incidental and incremental learning of vocabulary. *Computers & Education, 61*(1), 251–260. doi:10.1016/j.compedu.2012.10.006

Den, Y., Ogiso, T., Ogura, H., Yamada, A., Minematsu, N., Uchimoto, K., & Koiso, H. (2007). コーパス日本語学のための言語資源 [Linguistic resources for corpus Japanese]. *Journal of Japanese Language Science, 22*, 101–123.

Ehrich, J. F., Zhang, L. J., Mu, J. C., & Ehrich, L. C. (2013). Are alphabetic language-derived models of L2 reading relevant to L1 logographic background readers? *Language Awareness, 22*(1), 39–55. doi:10.1080/09658416.2011.644796

Gettys, S., Imhof, L. A., & Kautz, J. O. (2001). Computer-assisted reading: The effect of glossing format on reading comprehension and vocabulary retention, *Foreign Language Annals, 34*(2), 91–99. doi:10.1111/j.1944-9720.2001.tb02815.x

Grabe, W. (2004). Research on teaching reading. *Annual Review of Applied Linguistics, 24*(1), 44–69. doi:10.1017/S0267190504000030

Grabe, W., & Stoller, F. (2011). *Teaching and researching reading*. New York, NY: Longman.

Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Interaction analysis of a global online debate and the development of an interaction analysis model. *Journal of Educational Computing Research, 17*(4), 397–431. doi:10.2190/7MQV-X9UJ-C7Q3-NRAG

Hazelbeck, G., & Saito, H. (2012). A tool for creating glossed Japanese reading materials. *Information and Media Technologies, 7*(1), 318–327. doi:10.2197/ipsjjip.vol.20.1

Hori, K. [堀恵子] (2012). 文法項目データベース「はごろも」の海外における利用可能性をさぐる [Investigating possibilities of the use of the grammar database 'Hagoromo' overseas]. *Proceedings of the First National Conference of Japanese Dictionary Research*. Hagoromo Project Group. Retrieved from http://jisho.jpn.org/paper/hori_hagoromo.pdf

Horiba, Y. (2012). Word knowledge and its relation to text comprehension: A comparative study of Chinese- and Korean-speaking L2 learners and L1 speakers of Japanese. *The Modern Language Journal, 96*(1), 108–121. doi:10.1111/j.1540-4781.2012.01280.x

Horst, M. (2013). Mainstreaming second language vocabulary acquisition. *The Canadian Journal of Applied Linguistics, 16*(1), 171–188. Retrieved from https://journals.lib.unb.ca/index.php/CJAL/article/view/21299/24607

Hsu, C., & Yang, F. O. (2013). A vocabulary learning tool for L2 undergraduates reading science and technology textbooks. *International Journal of Science Education, 35*(7), 1110–1138. doi:10.1080/09500693.2013.764474

Hu Hsueh-Chao, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language,13*(1), 403–430. Retrieved from ERIC database. (EJ626518)

Jackendoff, R. (1997). *The architecture of language faculty*. Cambridge: MIT Press.

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning, 64*(1), 160–212. doi:10.1111/lang.12034

Jiang, X. (2011). The role of first language literacy and second language proficiency in second language reading comprehension. *The Reading Matrix, 11*(2), 177–190. Retrieved from http://www.readingmatrix.com/articles/april_2011/jiang.pdf

Kato, S. (2006). 日本語文法入門ハンドブック[*Handbook of Japanese grammar*]. Tokyo: Kenkyusha.

Koda, K. (1994). Second language reading research: Problems and possibilities. *Applied Psycholinguistics, 15*, 1–28. doi:10.1017/S0142716400006950

Koda, K. (1996). L2 word recognition research: A critical review. *Modern Language Journal, 80*, 450–460. doi:10.1111/j.1540-4781.1996.tb05465.x

Koda, K. (2002). Writing systems and learning to read in a second language. In W. Li, J. S. Gaffney, W. L. J. S., & Packard, J. L. (Eds.), *Chinese children's reading acquisition* (pp. 225–248). Boston, MA: Kluwer.

Komori, K., Mikuni, J., & Kondo, A. (2004). 文章理解を促進する語彙知識の量的側面 [Quantitative aspect of vocabulary knowledge promoting reading comprehension]. *Journal of Japanese Language Teaching, 120*, 83–92.

Kozawa, S., Uchimoto, K., & Den, Y. (2014). BCCWJ に基づく長単位解析ツール Comainu [Long-unit morphological analyser based on BCCWJ – Comainu]. *Proceedings of the 20th Conference of Natural Language Processing* (pp. 582–585). Retrieved from http://www.anlp.jp/proceedings/annual_meeting/2014/pdf_dir/P6-2.pdf

Kudo, T., & Matsumoto, Y. (2002). Japanese dependency analysis using cascaded chunking. *Proceedings of the 6th Conference on Natural Language Learning, 20*, 1–7. Stroudsburg, PA: Association for Computational Linguistics. doi:10.3115/1118853.1118869

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 20–34). Cambridge: Cambridge University Press.

Lee, K. (2002). Nominative case-marker deletion in spoken Japanese: an analysis from the perspective of information structure. *Journal of Pragmatics, 34*(6), 683–709. doi:10/1016/S0378-2166(02)00048-6

Lee, J. W., & Schaller, D. L. (1997). The relative contribution of L2 language proficiency and L1 reading ability to L2 reading performance: A test of the threshold hypothesis in an EFL context. *TESOL Quarterly, 31*(4), 713–739. doi:10.2307/3587757

Matsumoto, Y. (2007). Interaction of multiple factors in relative clause construal and acquisition. Commentary. *Studies in Second Language Acquisition, 29*(2), 367–374. doi:10.1017/S0272263107070234

McAlpine, J., & Myles, J. (2003). Capturing phraseology in an online dictionary for advanced users of English as a second language: a response to user needs. *System, 31*(1), 71–84. doi:10.1016/S0346-251X(02)00074-X

Myers, M. (1994). Various perspectives on educational linguistics gleaned from a collaborative project on the use of dictionaries. *Language Awareness, 3*(3-4), 193–200. doi:10.1080/09658416.1994.9959856

Nariyama, S. (2002). The WA/GA distinction and switch-reference for ellipted subject identification in Japanese complex sentences. *Studies in Language, 26*(2), 368–431. doi:10.1075/sl.26.2.07nar

Nation, L. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, L. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59–82. doi:10.1353/cml.2006.0049

Norman, D. A. (1988). *The psychology of everyday things*. New York, NY: Basic Books.

Ouyang, J. R. & Stanley N. (2014). Theories and research in educational technology and distance learning instruction through Blackboard. *Universal Journal of Educational Research, 2*(2), 161–172. doi:10.13189/ujer.2014.020208

Pulido, D. (2007). The effects of topic familiarity and passage sight vocabulary on L2 lexical inferencing and retention through reading. *Applied Linguistics*, *28*(1), 66–86. doi:10.1093/applin/aml049

Saville-Troike, M. (1984). What really matters in second language learning for academic achievement? *TESOL Quarterly*, *18*(2), 199–219. doi:10.2307/3586690

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal*, *95*(1), 26–43. doi:10.1111/j.1540-4781.2011.01146.x

Shiotsu, T., & Weir, C. J. (2007). The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. *Language Testing*, *24*(1), 99–128. doi:10.1177/0265532207071513

Shukatsu Navi. (2014, October 23). 面接のコツを知っていれば鬼に金棒です [Knowing some tips for interviews will give you a decisive advantage]. *Shukatsu Navi*. Retrieved from http://www.inautismcoalition.org/?p=80

Suzuki, T. (2013). 日本語学習者のための辞書使用のスキル養成のポイント [Essential points for enhancement of capability of dictionary use for Japanese language learners]. *Tokyo University of Foreign Studies Journal*, *86*, 131–158. Retrieved from http://repository.tufs.ac.jp/bitstream/10108/73470/2/acs086009_ful.pdf

Suzuki, Y. (2013, April 1). 見るオーディオ [Audio to look at]. *Music Bird*. Retrieved from http://musicbird.jp/audio_column/p5/

Tamaoka, K., & Yamada, H. (2000). The effects of stroke order and radicals on the knowledge of Japanese kanji orthography, phonology and semantics. *Psychologia: An International Journal of Psychology in the Orient*, *43*(3), 199–210. Retrieved from http://ir.lib.hiroshima-u.ac.jp/ja/list/department/46/p/3/item/14685

Tanabe, T., Takahashi, M., & Shudo, K. (2014). A lexicon of multiword expressions for linguistically precise, wide-coverage natural language processing. *Computer Speech & Language*, *28*(6), 1317–1339. doi:10.1016/j.csl.2013.09.001

Tateoka, Y. [舘岡洋子] (2005). ひとりで読むことからピア・リーディングへ：日本語学習者の読解過程と対話的協働学習 [From individual reading to peer reading: Reading process and interactive collaborative learning of Japanese language learners]. Tokyo: Tokai.

Toyoda, E. (2007). Enhancing autonomous L2 vocabulary learning focusing on the development of word-level processing skills. *The Reading Matrix*, *7*(3), 13–34. Retrieved from http://www.readingmatrix.com/articles/etsuko_toyoda/article.pdf

Toyoda, E., Arief, M. F., & Kano, C. (2013). Identifying useful phonetic components of kanji for learners of Japanese. *Japanese Language and Literature*, *47*(2), 235–272. Retrieved from https://www.aatj.org/journal-47-2

Yamashita, J. (2002). Mutual compensation between L1 reading ability and L2 language proficiency in L2 reading comprehension. *Journal of Research in Reading*, *25*(1), 81–95. doi:10.1111/1467-9817.00160

Yamauchi, H. (2008). 形態素解析に関する提案―日本語教育の視点から― [Suggestions for morphological analysis: From the perspective of teaching Japanese. *Jissen Japanese Literature*, *73*, 152–141. Retrieved from http://ci.nii.ac.jp/naid/110007056233

Yoshikoshi, K. (2012, September 28). 恐れと遠慮を捨てたとき人は真のリーダーになれる！[You can become a true leader when you let go of your fears and reserve]. *PHP Online Shuchi*. Retrieved from http://shuchi.php.co.jp/article/1160

Corresponding author: Etsuko Toyoda, etsuko@unimelb.edu.au

Please cite as: Toyoda, E. (2016). Evaluation of computerised reading-assistance systems for reading Japanese texts – from a linguistic point of view. *Australasian Journal of Educational Technology*, *32*(5), 94-107. https://doi.org/10.14742/ajet.2828