

画像認識技術を用いた体験型英単語学習支援システムの開発

馬 文鵬*, 皆月昭則**, 林 秀彦***, 伊藤陽介***

体験型英単語学習とは、従来の講義のような一方向の知識伝達型の学習ではなく、学習者が英単語の発見・理解・記憶を体験的に学習することを指す。本稿では、画像認識技術を活用した体験型英単語学習支援システムを提案し、その一部の機能を実装・評価する。提案したシステムはモバイルデバイスの最大の利点（モバイル環境と Web カメラ）を利用し、目の前の英単語を対象にし、認識・クラウド翻訳により単語情報を検索したり、学習者の問題を解決したりする特徴がある。本稿では、開発したシステムの性能評価とユーザビリティ評価実験について述べ、新たな英単語学習方法の可能性を示唆する。

[キーワード：体験型英単語学習，画像認識技術，学習支援システム，モバイルデバイス]

1. はじめに

現在、ポストPC時代と呼ばれ、いわゆる「モバイルデバイス時代」である。モバイルデバイスは、既存のPCに比べて携帯性に優れており、また無線通信によるインターネット接続を前提として設計されている。そのため、いつでも、どこでも、そのデバイスを利用できるという最大の利点がある。また、デバイスにカメラが搭載されている場合、これまでのPC時代に蓄積されてきた画像認識技術を応用して活用することができれば、人間の視覚システム機能を模倣したり、あるいは代替したり、さらには拡張することによって、人間のあらゆる活動をサポートする役割を果たすことができる。本稿では、これらのモバイルデバイスの利点に着目し、画像認識技術を用いた体験型英単語学習支援システムを提案し、そのシステムについて、一部の実装した機能を中心に述べる。本研究では、意味 (meaning)、スペリング (spelling)、発音 (pronunciation) の各方面から語彙力を高める英単語学習支援システムを提案する。提案したシステムはモバイルデバイスの利点を活用し、目の前の英単語対象を認識し、クラウド技術を通して、その単語の翻訳情報を学習者に提示する。この目の前の英単語対象とは、人の視覚認知機能に属する範囲である。例えば、本、写真、映像などに書かれている文字を指す。このシステムは、単語対象の認識機能・翻訳機能が日常生活または学習活動などのあらゆる活動を支援することができる。とくに、言語の習得、英単語による新事物の発見・認知などの活

動に良い影響を与えることをねらいとしたシステムである。

2. 提案システム

2.1 システム概要

開発環境はEclipse(4.2.1)+ADT(22.2.1)を用いた。また本システムはAndroid開発におけるVuforia SDK[1]を使用した。Vuforiaを使うところにより、タブレットが英単語を読み取り、クラウド翻訳サービスを通してスクリーンの上に翻訳文を提示する。

図1は本システムの実行画面である。本稿では、図に示すように英単語のスペリングと日本語の意味を提示する機能について述べる。図2に処理のフローチャートを表す。

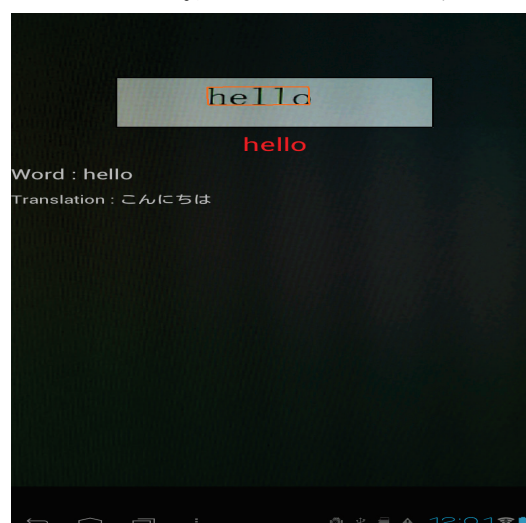


図1 実装したシステムの実行画面

* 鳴門教育大学 大学院(博士課程)教科・領域教育専攻 生活・健康系コース(技術・工業・情報)

** 釧路公立大学 経済学部

*** 鳴門教育大学 大学院 生活・健康系教育部

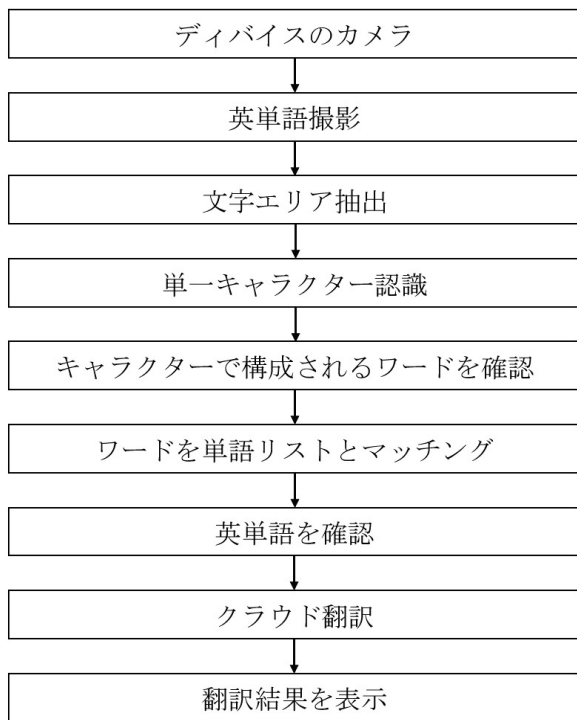


図2 システムのフローチャート

2.2 文字認識

画像認識による文字認識は英単語を対象にし、スマートフォンやタブレットのようなデバイスのカメラで単語を撮影し、画面中のワードを認識できる。実際に英単語を撮ると、オートズームにより連続的なフレームプレビューの中からフレームを抽出する。抽出されたフレームは画像として、メモリに保存する。システムはメモリにあるフレーム画像を取り込み、文字エリアを抽出する(図3)。システムは抽出された文字エリアの画像をOpenGLの関数によって二値化処理し、ワードをキャラクターで分割する。これらの単一キャラクターの画像がVuforiaに提供されたテンプレートマッチングにより、どのキャラクターかを確認する。VuforiaのテンプレートマッチングはUTF-8のキャラクターエンコーディングに依存し、

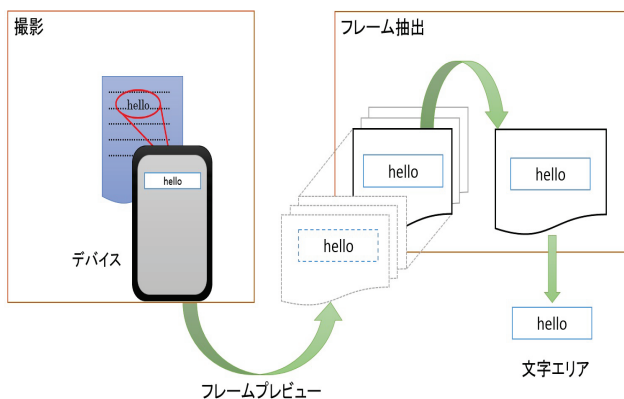


図3 フレームプレビューによる文字抽出

サポートするキャラクターは[LF]、[SPACE]、[']、[-]、[A]~[Z]、[a]~[z]の56個と限定している。これらのキャラクターで構成されるワードを受け入れることができる。そして、システムはこれらのワードをVuforia SDKに提供される13万超えの単語リスト(辞書)とマッチングする。マッチングが成功した場合は、そのワードが実在することになる。ワード認識の要件は以下のとおりである。

- ・文字数は2以上24以下である。
- ・ハイフンを含められる。(例えば、“mother-in-law”)
- ・スペースを1つ含められる。(例えば、“hot dog”)
- ・数字を含められない。

また、フォントとスタイルについて、システムはセリフ (serif) とサンセリフ (sans serif) のフォントを認識できる。そして、サポートするスタイルはプレーン (plain)、太字 (bold)、イタリック体 (italic)、アンダーライン (underline) を含め、4つの任意の組み合わせも含めることができる。様々なスタイルにも、同じ単語内の個々の文字に適用できる。すなわち、単一の単語は、異なるスタイルを含むことができる。とくに、システムは単語の文字が色で書かれる場合でも認識できる。

2.3 クラウド翻訳

図3の手法により認識した単語は、インターネットを通して、クラウド翻訳サービスのサポートにより単語を翻訳する。

翻訳のプロセスは図4に示している。システムは認識されたワードをURLでクラウド翻訳サービスに提出する。ワードが正しく翻訳された場合(成功の場合)、その結果のデータをJSONのフォーマットでシステムに返す。そのあと、システムはJSONのデータを解析し、翻訳された単語を抽出する。また、ワードが正しく翻訳されなかった場合(失敗の場合)、エラーを返す。

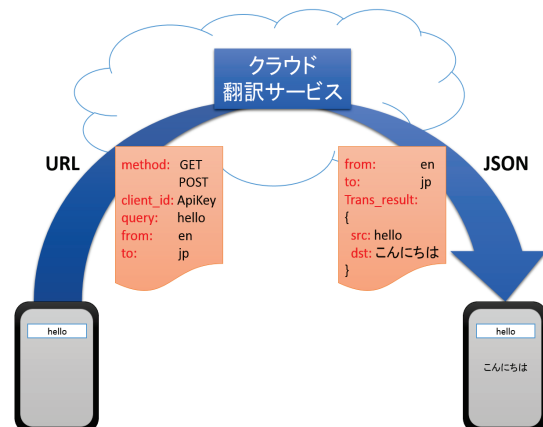


図4 クラウド翻訳サービスによる翻訳のプロセス

表1 各グループに分けた文字の認識率と翻訳の成功率

グループ (G)	G 1	G 2	G 3	G 4	G 5	G 6
実験単語 (10項目/G)	with for through might any used to about favor bring from	so faith around take leave whole may celebrate how weird	inquiry something still as right to hold available would look forward to	confidence since order love division awesome ensure serious implement on behalf of	nothing across behind what remove evidence snow get into I'm afraid	issue statement would like have appreciate circumstance must follow reputation fall
認識率	100%	100%	90%	80%	90%	100%
翻訳 成功率	100%	100%	90%	80%	90%	100%

3. 評価実験

提案システムは、デバイスのカメラで文字を読み取って、クラウド翻訳サービスを通して日本語に翻訳する機能である。実装した機能が正しく動作することを確認する性能評価実験と、実装した機能が既存の英単語情報検索手段と比較して、学習者にとってどのような有用性を見出すかを評価するユーザビリティ評価実験を実施した。また、ユーザビリティ評価実験では今後のシステム改良のためのユーザビリティ評価データを収集することを目的とした。以下に、評価実験について述べる。

3.1 性能評価実験

(1) 目的

性能評価実験は、瞬時翻訳の機能を実装し、機能が正しく実装されていることを確認するための性能評価実験を目的とした。とくに、文字（英語）の認識率、翻訳の成功率、認識してから翻訳するまでの平均処理時間について性能を評価した。

(2) 方法

Web上のgoo辞書[2]に提供される検索ランキングとして表示された60個の英語（単語またはフレーズ）を対象とする。60個の英語は、表1に示すように、6つのグループに分かれ、1つのグループは10項目で構成されており、単語やフレーズが含まれている。このリストを対象として、文字認識率と翻訳成功率を比較する。ここで、文字認識率は、1グループ10項目に対して、カメラが正しく認識した項目の割合である。翻訳認識率は、各項目に対して、正しく翻訳した項目の割合である。

(3) 結果

表1は各グループに分けた文字の認識率と翻訳の成功率の結果である。認識率はグループ4が最低(80%)となり、グループ1、グループ2、グループ6が最高(100%)となった。また、表1に示すように、各グループの翻訳の認識率は、各グループの文字の認識率にそれぞれ対応した結果となった。

(4) 考察

実験の結果、グループ3の「look forward to」、グループ4の「on behalf of」と「implement」、グループ5の「I'm afraid」4つの英語による文字認識が失敗した。その原因を次に分析する。

原因1：項目「look forward to」はシステムに「look」、「forward」、「to」という3つの文字として認識された。これは、それぞれの単語としての認識は成功しているが、項目としての認識は正しくない。このように認識された原因は、「スペースを1つのみ含められる」という要件を満たしていないためと考察できる。

原因2：項目「on behalf of」は上記と同じ理由だと考察できる。

原因3：項目「implement」は「lenient」と認識された。しかしながら、「implement」の文字サイズを大きく表示した場合は、正しく「implement」として認識された。また、被験者がデバイスを持つ姿勢を調整した場合は、「implement」として正しく認識された。

表2 電子辞書と提案システムによる実験結果

英単語 検索手段	グループ						総時間	平均 時間	
	1	2	3	4	5	6			
電子辞書	XD (A)	45.25s	47.80s	55.75s	47.75s	42.51s	45.56s	284.62	4.74s
	XD (B)	50.03s	44.33s	60.93s	43.08s	39.81s	35.15s	273.33s	4.56s
	XD (A+B)/2	47.64s	46.07s	58.34s	45.42s	41.16s	40.36s	278.98s	4.65s
システム	AT700 (A)	9.28s	11.18s	10.59s	9.27s	9.96s	10.66s	60.94s	1.02s
	AT700 (B)	9.76s	10.55s	9.87s	10.69s	10.31s	9.96s	61.14s	1.02s
	AT700 (A+B)/2	9.52s	10.87s	10.23s	9.98s	10.14s	10.31s	61.05s	1.02s

原因4：項目「I'm afraid」は「I'm」と「afraid」という2つの文字として認識された。アポストロフィ「'」は、サポートされているキャラクターであり、「I'm afraid」は2.2節に記した要件「スペースを1つ含まれる」にも準拠している。しかし「I'm afraid」は英語のフレーズに属するため、フレーズは、このシステムの対象外であるので、正しく認識されなかったと考察している。

翻訳成功率については、認識された文字はすべて正しく翻訳された。そのため、翻訳成功率は文字の認識率に対応した。文字認識と翻訳の平均処理時間は、カメラに画像が表示されてから1秒以内であった。

この翻訳の実験では、認識率が文字の構成（フレーズかどうか）、文字のサイズ、ユーザの姿勢等に影響され、翻訳の成功率は文字の認識率に依存している。そのため、提案システムは、文字サイズおよび被験者がデバイスを持つ姿勢等のカメラと対象の位置関係について、どのような制限があるのかを今後明らかにすることによって、利用の可能性を広げることが期待できる。

3.2 ユーザビリティ評価実験

(1) 目的

既存の英単語検索手段と比較して、提案システムは検索時間が短いという仮説を立て、ユーザビリティ評価実験では、この仮説を検証することを目的とした。また同時に今後のシステム改良のためのユーザビリティ評価データを収集することも合わせて行った。

(2) 方法

この実験では、2人の被験者が60個の英語（単語またはフレーズ）を電子辞書（CASIO Ex-word XD-SF4800[3]）に入力し、翻訳情報が表示されるまでの時間（総時間）を計算する。また、電子辞書による入力と瞬時翻訳機能の時間効率を比較するため、1つの単語あたりの時間（平均時間）を計算する。時間の測定は、60個の英語（単語またはフレーズ）を6つのグループに分けて行った。

(3) 結果

実験結果を表2に示す。表2に示すXD(A)、XD(B)、XD(A+B)/2は、それぞれ被験者A、被験者B、そして被験者Aと被験者Bの平均の電子辞書による入力の結果を示し、AT700(A)、AT700(B)、AT700(A+B)/2は提案するシステムの翻訳機能を使用した結果を同様に示している。

表2に示すように、電子辞書による実験結果は、Aの総時間284.62秒となり、平均時間が4.74秒となった。Bの総時間273.33秒となり、平均時間が4.56秒となった。(A+B)/2の総時間278.98秒となり、平均時間が4.65秒となった。提案システムの翻訳の実験結果は、総時間が61.05秒となり、平均時間が1.02秒となった。総時間の比較結果から、システムの翻訳機能は、被験者Aと被験者Bの平均による電子辞書の入力に比べて、60個の英語に対して約218秒分の短い時間で検索できることを示した。また、これは時間効率に換算して、約4.56倍の高い効率であることを示した。

(4) 考察

表2に示した平均時間の結果から、既存の英単語検索手段と比較して、提案システムは検索時間が短いことが示された。

電子辞書の場合、とくに、実際に実験したとき、英語（単語またはフレーズ）の文字数が少ない場合（例えば：「as」）、入力する時間が約1秒であったり、文字数が多い場合（例えば：「construction」）、十数秒であったり、文字数によって時間にばらつきが生じる結果となった。これに対して、瞬時翻訳の場合は1秒以内であったので、ばらつきが少ない。

goo辞書による英語は単語とフレーズを含め、単語の場合は、電子辞書とデバイスに入力して直接翻訳できる。フレーズの場合は、電子辞書とシステムのどちらも認識できない状況がある。この状況に対して、電子辞書が認識できなくても、提案システムは認識できない単語を追加リストに加え、認識できるようにする機能を備えている利点がある。また、電子辞書への文字入力は単語の文

字数によって影響を受けるが、提案したシステムは、その点は大きく影響しないため、単語の文字数を多くした場合は、電子辞書に比べて時間効率はさらに高くなることが予想できる。

4. 考察

提案した英単語学習支援システムを実装し、そして文字（英語）の認識率、翻訳の成功率、認識してから翻訳するまでの平均処理時間について評価実験を行った。システムの翻訳機能における評価実験では、Webカメラを用いた入力インターフェースから獲得した情報を画像認識し、文字情報をカメラで読み取って瞬時翻訳できる結果が得られた。電子辞書、検索エンジンなどの一般的な翻訳手段と比べて、提案システムの翻訳機能は、キーボードからの入力がいらなく、デバイスを目の前の対象に向け、撮影することで情報を入力できる。そのため、ユーザはモバイルデバイスのキーボードからの入力ミスによる煩わしさから解放される。また、目の前の対象を撮影できるため、提案したシステムは実世界の情報を検索できる特徴を有している。従って、言語学習などの場面における活用の効果を高めることが期待できる。

今後の課題として、本システムは、英単語を学習することを支援するシステムであり、体験を通した利用に重点が置かれているため、システム開発時点においては利用者の行動予測モデルの確立は十分ではない点が挙げられる。そのため、システムの利用テスト期間を経て、今後は利用者に対するシステム利用上の注意説明を行うための資料準備とそれに伴う問題点の抽出を予定している。システム利用における問題点の抽出には、クラウド翻訳サービスを利用する点についての利用者への配慮事項の追記説明等が挙げられる。

5. まとめ

言語能力の育成は重要な教育の課題である。言語の重要性は母国語教育などの限られた場面にとどまらず、学習者の生活全般あるいは生涯にわたっている。その最適な学習時期については本稿の言及する範囲ではないが、学校への入学は、ひとつの大きな節目であり、この時期を目途に一定の言語能力の習得が期待される。

現在の日本では、言語教育は主に学校・家庭において育成されている。具体的には、日常の遊びの中の活動あるいは通常のコミュニケーション活動を通して、教科書などの言語教材その他の教材に親しませることに中心がおかれている。提案する英単語学習支援システムは、従来の言語教育環境に依存しない。すなわち、学校・家庭の環境ではない日常生活において、デバイスの画像認識技術により、学習者にとって新事物を認識し、その英

語対象の基本情報を示すことができる。提案したシステムは、文字情報に対して翻訳する機能を備えているため、学習者は第二言語の学習にも役立てることができる。とくに、学習者が実際の生活環境の中からシステムを介して言語を発見することの喜びや、言語学習活動を通じた体験による発見・感動を導き、実世界における言語の理解を深めることに波及する。また、提案したシステムは、自然に遊びのなかに取込まれるようにして、学習者がデバイスを遊具として活用するなかで言語学習に興味を覚え、言語能力を育成することも期待している。

参考文献

- [1] Vuforia SDK, <http://www.vuforia.com/>
- [2] goo 辞書 検索ランキング, <http://dictionary.goo.ne.jp/ej/>
- [3] CASIO Ex-word XD-SF4800 の仕様, http://casio.jp/exword/products/model/spec/?code=XD-SF4800&navi=11D_01