

漢字熟語の類推を支援する辞書システム“KIDS-II”の構築

越 智 洋 司[†] 矢 野 米 雄[†] 林 敏 浩^{††}

外国人のための漢字熟語学習においては、単に暗記するのではなく、熟語の構成を学習者に意識させて熟語を教え、表記面からの類推力を獲得させる方法が提案されている。しかし、漢字は複数の意味、読みを持つため、学習者は誤った類推を行う可能性がある。このとき、学習者が解答した熟語を正誤判定するだけでなく、学習者が適用した知識の正当性を分析、説明する必要がある。本論文では、熟語の構成パターンを類推するためのルールとして用い、そのルールによって類推候補を動的に扱う漢字熟語辞書 KIDS-II を提案する。KIDS-II は、任意の漢字熟語に対する意味、読みの類推候補を動的に導出し、知識ベースに収録された漢字熟語データを基にした使用頻度によって、類推候補の正当性を評価する機能を持つ。また、熟語の読みの変化の認識や、熟語知識ベースに未登録の読みや熟語から登録熟語を検索できる。KIDS-II を用いることで、学習者は自身の漢字熟語に対する意味、読みの類推を評価できる。

KIDS-II: A Dictionary System for Supporting Analogy of Kanji Compounds

YOUJI OCHI,[†] YONEO YANO[†] and TOSHIHIRO HAYASHI^{††}

A method for teaching Japanese kanji compounds to foreigners is proposed, which is not just learning kanji compounds by heart, but learning consciously the construction of them. In this way, learners could better acquire the ability of analogy. However, they can make an incorrect analogy, because each kanji usually has more than one meaning and reading. A teacher must analyze learner's applied knowledge and explain them their mistakes. We developed a kanji compounds dictionary KIDS-II which treats the analogy candidate dynamically using the composition pattern of the kanji compounds. KIDS-II has a function which evaluates the analogy candidate of the kanji compounds meaning and reading by the frequency of its use. Our system can display the reading changes in the compounds, and it also can derive the appropriate compounds for the readings or compounds non-stored in the knowledge base. Using our system KIDS-II, the learners can judge the correction of their own answers.

1. はじめに

近年、来日する外国人の増加により、日本語教育環境の整備が早急に望まれている。しかし、外国人学習者に対する日本語教師の不足により、その解決策として CAI (Computer Assisted Instruction) に期待が寄せられている¹⁾。日本語の学習において漢字学習は避けられない。特に、日本語には 2 字以上の熟語が多く使用されるため、漢字熟語学習は欠かせない。本論文では、漢字熟語の基本となる二字熟語（以下熟語と記述）を対象領域とする。

日本人は小学校から国語教育で熟語を覚えると同時

に、文脈に適した熟語の選択や、未学習の熟語に対応する能力を身につけている。外国人の熟語学習においても、単に熟語の意味や読みを暗記するのではなく、個々の漢字の意味と読みを学習者に意識させた熟語の教授や、表記面からの意味と読みの類推力の獲得支援が提案されている^{2)~5)}。しかし、漢字は複数の品詞や読みを持つため、学習者が誤った類推を行う可能性がある。たとえば、「じんこう」を「人校」、「人情」を「ひとじょう」のように、同音漢字や音訓の使い分けが誤った類推や、「出発」を「しゅつはつ」のように、読みの変化に未対応の場合等があげられる。また、「長所」と「所長」のように、順序によってまったく意味が異なる語を正しく理解できない場合もある。この場合、教師は学習者の解答した熟語を単に正誤判定するだけでなく、適用した知識の正当性を分析し、正しい知識を教授する必要がある。

[†] 徳島大学工学部

Faculty of Engineering, Tokushima University

^{††} 佐賀大学理工学部

Faculty of Science and Engineering, Saga University

言語学、国語学の分野では、熟語の構成に関する研究がいくつかされており、熟語の意味構成の典型的なパターンが存在が確認されている⁶⁾。文献7)は、「人間はこのパターンを規則として意識的に利用するのではなく、用例をあげる形で利用する」ことを観測している。そこで、我々は熟語学習においても、用例と関連づけながら類推能力を獲得させることが重要と考える。しかし、熟語ごとにそのパターンは異なり、例外も存在するため、学習者が行った様々な類推パターンに対する適切な説明や用例の提示を行うことは教師にとって負担である。一方、計算機はこの種の大量のデータを扱ったパターンマッチングを得意とする。そこで本論文では、計算機による支援が適切と考え、学習者の類推を支援する電子辞書の枠組みを採用する。

熟語を扱った電子辞書システムの研究に、CASTEL/J⁸⁾、KIDS⁹⁾がある。CASTEL/Jは、日本語教育用教材データとして、漢字、単語、用例等を持つ辞書データベースである。KIDSは、熟語の意味構成を関連知識として持つ辞書システムである。これらのシステムは、読み、意味、用例といった一般に使われている熟語の知識が登録されているが、類推情報を扱っていない。そのため、システムは知識ベースに収録されていない学習者の類推を、誤りまたは未登録語としか扱えない。熟語の類推を扱う電子辞書構築において、類推候補をすべて登録する枠組みを採用すると登録語数は際限なく増加し、構築が困難となる。また、計算機における未登録語の処理に関する研究では、既知の単語(用例)を用いて未知熟語の意味を推定するシステムが開発されている¹⁰⁾。しかし、自然言語処理における未知語獲得を目的とするため、教育システムとしてそのまま適用できない。

そこで我々は、常用漢字1945字と熟語の構成パターンを用いて意味と読みの類推候補を動的に生成し、知識ベースに登録された熟語との関連づけにより、学習者の類推を支援する漢字熟語辞書KIDS-II (Kanji Idiom Dictionary System-Idiom Inference version)¹¹⁾を構築した。本論文では、2章で熟語の類推規則を説明し、3章でKIDS-IIの概要を述べる。また、4章、5章、6章でKIDS-IIの知識表現、検索・類推手法ならびに、ユーザインタフェースの詳細を述べ、7章で本システムの実験的評価およびCAIシステムへの応用について述べる。

2. 漢字熟語の類推規則

日本人は、未知の熟語に遭遇しても、構成する漢字が既知であればその意味や読みを類推し、既知の漢字

の知識を利用して、文脈に適した熟語を類推できる。文献7)は、人間が未知熟語の意味を推定する際に、(i)複合語の構成規則、(ii)構成要素である漢字の原型の意味、(iii)語スキーマ、(iv)類似する単語の構成についての知識、(v)外界に関する知識を用いることを解明している。本研究では、熟語の構成規則を意味規則と読み規則に分け、類推規則として扱う。本章では、各規則の詳細について述べる。

2.1 意味規則

国語学の分野では熟語の意味構成に関する典型パターンが確認されており、文献6)は熟語を構成する漢字の意味の係り方を「主述」、「修飾」、「並列」、「補足」、「認定」の5つの基本構造に分けている。また、文献2)は、日本語教育の立場から漢字の品詞を「名詞」、「動詞」、「形容詞」の3つに分け、基本構造を構成する漢字の品詞の並び方によって表現している。しかし、この手法は「認定」の関係を考慮していない。一方、文献12)は、計算機による二字漢字の未登録語処理において、必要な漢字の形態素と二字熟語を構成する漢字間の関係を分類している。そこで本論文では、「名詞」、「動詞」、「形容詞」を基に、表1に示す12種の品詞を用い、熟語の品詞構造を表現する。さらに品詞構造を学習者が理解しやすいように、

- (1) 修飾の関係(後部が修飾語)
- (2) 修飾の関係(前部が修飾語)
- (3) 主述の関係
- (4) 並列関係

の4パターンに分類した。本論文では、この4パターンで表される品詞構造を意味規則とする。表2に、意味規則とその例を示す。なお、表1中の形容詞的名詞とは「名詞」+「の」を示し、前・後置語とは、「不」、「非」、「否」のように、名詞、動詞、形容詞以外の品詞で、熟語中の位置に依存する語を示す。また、規則中の「*」はすべての品詞パターンを示す。

2.2 読み規則

漢字は音訓の読みを持つことから、熟語の読みには次の4パターンの組合せが存在する。

表1 漢字の品詞
Table 1 Part of speech of kanji.

品詞	記号	説明	記号
名詞	No	形容詞的名詞	Na
自動詞	Vi	他動詞	Vt
形容詞的動詞	Va		
形容詞(述語)	Ap	形容詞(限定)	At
副詞	Ad		
数詞	Nu	助数詞	Nx
前置語	Pr	後置語	Po

表2 意味規則
Table 2 Meaning rules.

関係	品詞構造	例
修飾の関係 (後部が修飾語)	Vt + No	読書
	Vi + Ad	革新
	Pr + *	不動
修飾の関係 (前部が修飾語)	No + Vi	期待
	Ad + Vi	細分
	Ad + Ap	極大
	At + No	難問
	Na + No	牛乳
	Va + No	課題
	Nu + Nx	一個
	* + Po	安否
主述の関係	No + Vi	地震
	No + Ap	胃弱
並列の関係	At + At	重大
	Vi + Vi	勝負
	No + No	子孫

表3 読みの変化規則の例
Table 3 Example of reading rules.

変化の種類	パターン	例
促音	く + こ	学校(がっこう)
	ち + こ	日光(にっこう)
濁音	お + そ	大空(おおぞら)
半濁音	ん + は	乾杯(かんぱい)

- (1) 音+音(漢語読み): 水準, 国家, 漢字
- (2) 訓+訓(和語読み): 左手, 花見, 大空
- (3) 音+訓(重箱読み): 本箱, 新型, 派手
- (4) 訓+音(湯桶読み): 雨具, 相性, 湯気

本論文ではこの4パターンを読み規則の基本とする。また、熟語には「学校(がっこう)」、「大空(おおぞら)」のような促音や濁音、「大和(やまと)」、「天王(てんのう)」のような熟字訓や連声の熟語が存在する。これらの語は、個々の漢字の読みが変化している。文献13)は、促音が発生する規則として、音節のつながりのパターンをあげている。そこで本論文では、前後の漢字の読みのつながり方に着目し、促音だけでなく、濁音や半濁音についても同様に音節パターンを抽出し、読み変化規則とする。表3に読みの変化規則の例を示す。

3. KIDS-II

我々は、2章の類推規則に基づいて類推候補を動的に扱い、熟語学習における学習者の類推を支援する電子辞書システム“KIDS-II”を構築した。KIDS-IIは、学習者が未習の熟語に遭遇した場合に、意味や読みに関して行った類推の正当性の評価と、誤った類推からの正しい知識の検索を支援する。本章では、KIDS-II

の特徴とシステム構成について述べる。

3.1 特徴

KIDS-IIは、熟語学習における学習者の類推を支援する電子辞書として、以下の特徴を持つ。

- (1) 単漢字と関連づけた熟語検索
- (2) 類推候補の導出と評価
- (3) 類推候補に関連する熟語の導出

(1)は、熟語中での各漢字の意味や読みの情報を動的に解析し、その結果を提示することで単漢字の知識に着目した熟語の知識獲得を支援する。また読みに関しては、促音、濁音、半濁音、特殊音(熟字訓・連声等)を含んだ熟語についても、正しく読みを分割する。たとえば、「学校」の読みは「がっ」「こう」となる。

(2)は、2章の意味規則と読み規則により、意味と読みに関する類推候補を導出する。しかし、類推候補には熟語としてほとんど使われないような読みや意味を含んだものが存在する。また、読み変化規則も絶対的なものではなく、変化しないケースも存在する。たとえば、熟語の読みはほとんどが漢語読みであるため、「牛」という漢字は「ぎゅう」と読むのが一般であるが、「子牛」のように、訓が使われる場合もある。我々は、学習者の行った誤った類推が事例として存在するか、頻繁に起こるかといった評価が行える環境も必要と考える。そこでKIDS-IIは、知識ベースの熟語データ(2832語)を基にした漢字の読みと品詞の出現回数を利用する。学習者の誤った類推パターンが、熟語データ中で利用されている割合を示す頻度情報を算出し、類推候補に付加する。この頻度情報により類推候補を順位づけて提示することで、KIDS-IIは学習者に自らの行った類推の評価を支援する。頻度情報の算出方法については、5.3.4項で述べる。

(3)は、学習者の誤った類推から知識ベースに未登録の熟語や読みが入力された場合、未登録語の意味と読みを類推し、該当する熟語を導出する。そして学習者の正しい知識の獲得および、関連語の提示による体系的な熟語の学習を支援する。たとえば、学習者が「人木」、「任気」という類推結果を入力した場合でも本システムは「人気」を導出できる。

なお、KIDS-IIにおける未登録語とは「人木」、「任気」といった、実在せず、しかも知識ベースに登録されていないものを指す。しかし、熟語は数が多く、新語も存在するため、大規模な知識ベースを用意しても、実際に存在するが知識ベースに登録されていない熟語(未収録熟語)をどう扱うかが問題となる。つまり、学習者が知識ベースに未登録な語を入力した場合、その語が実在する熟語であるかの区別が必要となる。この

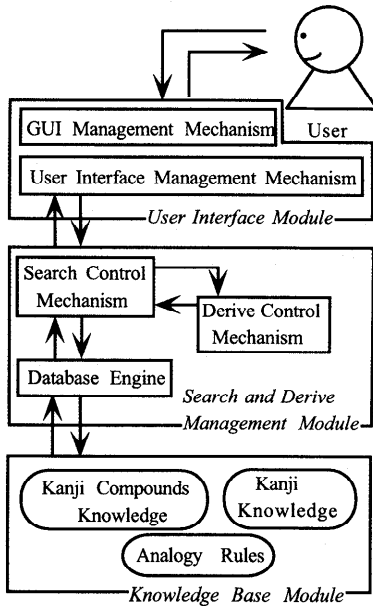


図1 システム構成

Fig. 1 System configuration.

問題に対し、本論文では KIDS-II の知識ベースは教師の教育方針が反映されているとする。そこで、未収録熟語については、教師の教育方針の範疇を超えたものとし、KIDS-II ではこの区別を扱わない。

3.2 システム構成

KIDS-II は、類推規則と単漢字の知識により、熟語の類推候補を扱う枠組みを採用する。本システムは、知識ベース部、検索・導出管理部、ユーザインタフェース部から構成される。システム構成を図1に示す。

(1) 知識ベース部 知識ベース部は、任意の熟語に関する意味と読みの類推候補群を導出するための類推規則と単漢字の知識を持つ。しかし、類推規則と単漢字の知識のみでは、実際に用いられる熟語の意味と読みを決定することは困難である。そこで、学習者に正しい知識を教授するための熟語知識もあわせて持つ。また、類推候補を頻度によって評価するために、知識ベースに収録された熟語中での品詞と読みの出現回数を単漢字知識に記述する。この出現回数は、熟語知識として収録された2832語の熟語を基にしている。読みについては、熟語中の漢字の位置とペアとなる読みの種類(音, 訓)に着目し、熟語中での出現回数を記述する。読み変化規則についても、促音, 濁音, 半濁音が発生した回数を記述する。同様に、意味についても熟語中の漢字の位置に着目した品詞の出現回数を記述する。各知識はリレーショナルデータベースで管理する。知識ベースの詳細は4章で述べる。

(2) 検索・導出管理部 検索・導出管理部は、検索制御機構、導出制御機構とデータベースエンジンより構成される。検索制御機構は、知識ベースから漢字と熟語の意味や読み等を検索するための関数群で構成される。各関数は、検索のためのSQL文を動的に生成し、データベースエンジンに伝達し、検索結果を受け取る。導出制御機構は熟語・漢字知識を利用して(i)熟語読みの解析,(ii)意味, 読みの類推候補の導出,(iii)頻度情報の算出を行う。詳細は5章で述べる。

(3) ユーザインタフェース部 ユーザインタフェース部はGUI管理機構と対話管理機構から構成する。前者は、グラフィカルユーザインタフェースを中心とした操作環境を用意し、システムとユーザ間の対話を行う。後者は、ユーザのリクエストを検索・導出管理部へ伝達し、検索・導出結果をインタフェース管理機構へ送る。詳細は6章で述べる。

4. 知識ベース部

KIDS-II は図2のように、漢字知識と熟語知識、類推規則で構成する知識ベースを持つ。本章では各知識の知識表現について述べる。

4.1 漢字知識表現

漢字知識は、常用漢字1945字を対象とし、“読み知識”、“品詞知識”、“意味-品詞知識”のテーブルを持つ。“読み知識”には各漢字の音読みと訓読み、および各読みの出現回数を記述し、3446組のデータを収録した。出現回数は“前音”、“前訓”、“後音”、“後訓”のフィールドで表現する。“前音”にはその漢字の読みの前部に音読みが来る回数、“前訓”には訓読みが来る回数を記述する。同様に、“後音”にはその漢字の読みの後部に音読みが続く回数、“後訓”には訓読みが続く回数を記述する。たとえば、「左手」の場合、「手」の前に「ひだり」という訓読みが来るため、「手(て)」の“前訓”がカウントされる。また、「左」の後ろに「て」という訓読みが続くので、「左(ひだり)」の“後訓”がカウントされる。“品詞知識”には、熟語中で用いられた品詞の位置に基いて、出現回数を“前部”、“後部”のフィールドに記述し、5669組のデータを収録した。“前部”は、該当する品詞が熟語の前部で用いられた場合を示す。“意味-品詞知識”には各漢字の意味と対応する品詞を記述し、6613組のデータを収録した。なお、意味については、同義語検索が行えるように簡潔な語彙で表現した。

4.2 熟語の知識表現

熟語知識は、2832語の二字漢字熟語を対象とし、“意味知識”、“読み知識”のテーブルを持つ。“意味知識”

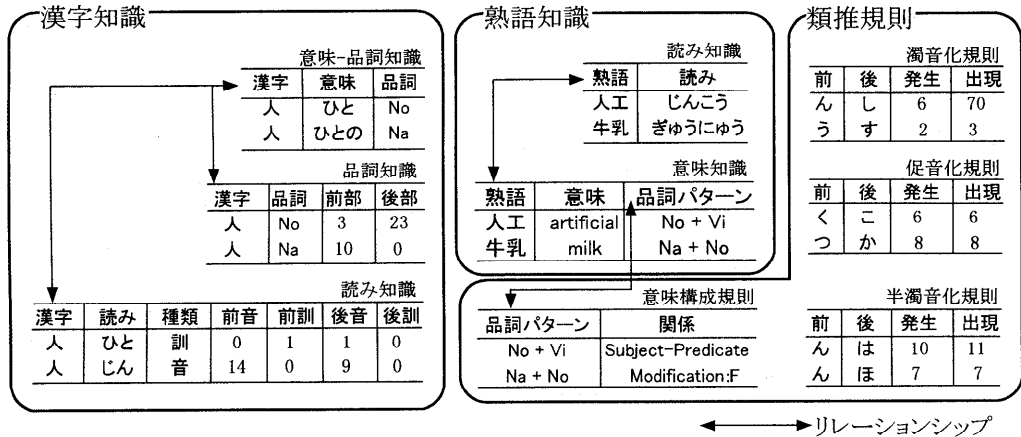


図2 知識表現
Fig. 2 Knowledge representation.

には、学習者が理解するための各熟語に対する意味と、熟語に対応する品詞パターンを記述する。品詞パターンは、2.1 節で述べた意味規則の中で該当する品詞パターンであり、熟語の正しい意味規則の導出に用いる。“読み知識”には、熟語に対応する読みを記述する。“意味知識”、“読み知識”ともに 2832 組のデータを収録した。

4.3 類推規則の知識表現

類推規則は、読み規則として“促音化規則”、“濁音化規則”、“半濁音化規則”のテーブルと、意味規則として“意味構成規則”のテーブルを持つ。読み規則の各テーブルには、熟語知識から抽出した促音、濁音、半濁音の発生した読みの連結パターンと、その出現回数、および読みの変化が発生した回数を記述する。

“意味構成規則”には、2.1 節で述べた意味規則の品詞パターンに対応する意味関係名を“Subject-Predicate”（主述），“Modification:F”（修飾：前部が修飾語），“Modification:B”（修飾：後部が修飾語），“Parallel”（並列）で記述する。本知識ベースには、漢字熟語 2832 語を基に促音化規則 32 組、濁音化規則 49 組、半濁音化規則 10 組を収録した。

5. 検索・導出管理部

本モジュールは、知識ベースに収録した類推規則と漢字知識、熟語知識を制御し、(1) 漢字、熟語知識の検索、(2) 熟語の読み解析、(3) 意味と読みの類推候補の導出、(4) 頻度情報の算出を行う。

5.1 漢字・熟語知識の検索

検索制御機構は、単漢字および熟語に対して、i) 読みからの知識検索、ii) 漢字からの知識検索、iii) 類推規則の検索を可能とする知識検索関数群を持つ。各

関数はインターフェースからの要求に対応する SQL を自動生成し、データベースエンジンに送り、結果を受け取る。これらの関数を組み合わせ、知識ベースの検索を実現する。

5.2 熟語の読み解析

導出制御機構は、漢字の“読み知識”と“類推規則”を用い、任意の熟語の (1) 読みの分割情報、(2) 音訓情報、(3) 読みの変化情報の解析を行う。(1) は、まず単漢字の読みを組み合わせ、読み候補を導出する（以下：読み候補 α 群）。そして、読み候補 α 群と熟語の読みが一致した場合の組合せから、読みの分割を決定する。(2) は、(1) で分けられた読みに対して、漢字知識ベースを検索し、音訓情報を抽出する。(3) は、読み候補 α 群と熟語知識ベースに記述された実際の熟語の読みを比較し、一致しない場合に熟語の読みが変化したと判断する。そして読み変化規則に対応する音節パターンの熟語については、読みを変化させる（以下：読み候補 β 群）。最後に、読み候補 β 群と熟語の読みを比較することにより、熟語の読みが促音、濁音、半濁音化しているか、特殊音であるかを解析する。図3に読みの解析手順を示す。

5.3 類推候補の導出

導出制御機構は、読み、意味、熟語候補と各候補の頻度情報を導出する。本節では、その手法について述べる。

5.3.1 読み候補の導出

読み候補の導出は、5.2 節で述べた読み候補 α 群と読み候補 β 群を利用する。また、前部、後部の読みの音訓情報から出現回数を検索し、読み候補 α 群、読み候補 β 群に対して頻度情報を付加する。

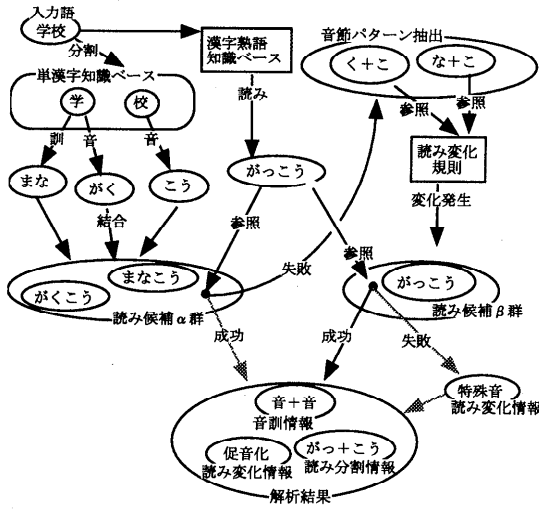


図3 読み解析の手順
Fig. 3 The process of the reading analysis.

5.3.2 意味候補の導出

意味候補の導出は、まず、各漢字の品詞を組み合わせ、2.1節で述べた意味規則と一致する品詞パターンを決定する。そして、その品詞に対応する各漢字の意味を検索し、組み合わせることで意味候補を導出する。さらに品詞の位置に関する出現回数から、導出された意味候補に対し、頻度情報を付加する。

5.3.3 熟語候補の導出

熟語候補の導出は、(1)未登録読みからの導出、(2)未登録熟語からの導出に分けられる。

(1)未登録読みからの熟語導出 未登録読みからの導出は、通常熟語導出と、読み変化規則をともなった熟語導出に分けられる。通常は、入力された読みの文字列を、先頭文字から順に区切ることで前部と後部に分け、各々の漢字検索が成功した分割パターンを読み分割候補とする。そして、各読みに対応する漢字を組み合わせ、漢字熟語知識ベースに存在する漢字列を熟語候補として出力する。しかし、促音、濁音等の読みの変化をともなった読みが入力された場合、分割した各読みに対応する漢字を検索できない。そこで、読み変化規則を用いて読みの原形候補を生成し、熟語候補の導出を行う。図4に読みからの熟語候補の導出手順を示す。

(2)未登録熟語からの熟語導出 未登録熟語からの熟語導出は、読みに着目する場合と意味に着目する場合がある。前者は、5.2節の熟語の読み解析手順と同様に、入力された漢字列から熟語読み候補α群とβ群を導出する。そして、それらを漢字熟語知識ベースで検索し、読み候補に対応する熟語を熟語候補として導

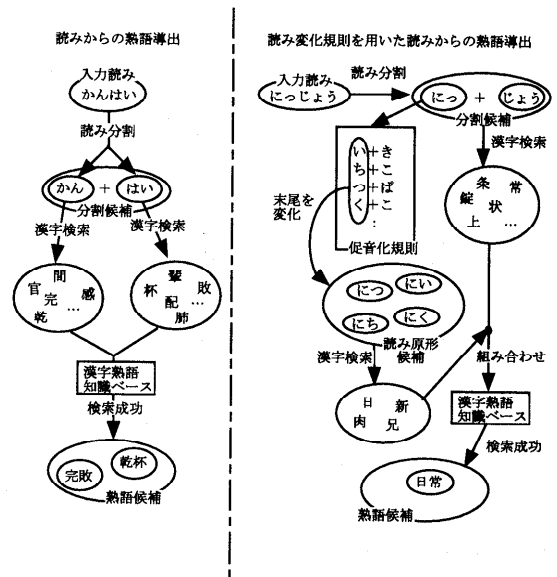


図4 熟語候補の導出手順(1)
Fig. 4 The process of deriving compounds (1).

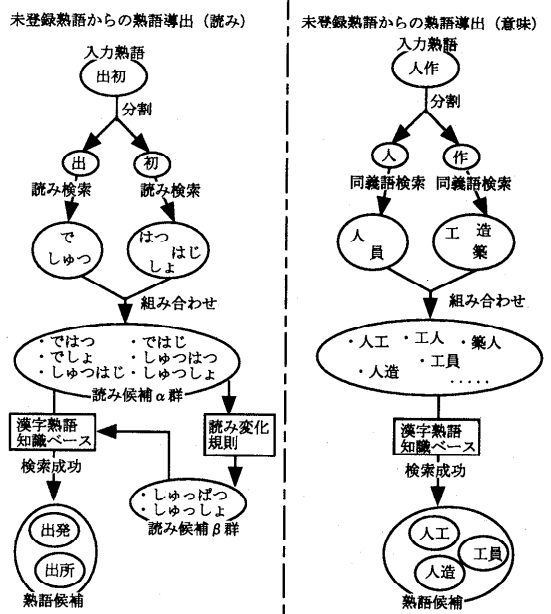


図5 熟語候補の導出手順(2)
Fig. 5 The process of deriving compounds (2).

出する。後者は、前部、後部の漢字と同じ意味を持つ漢字を検索し、その組合せから、漢字熟語知識ベースに収録されている熟語を検索する。図5に未登録熟語からの熟語候補の導出手順を示す。

5.3.4 頻度情報の算出

KIDS-IIは、導出する意味と読みの各候補に対し、

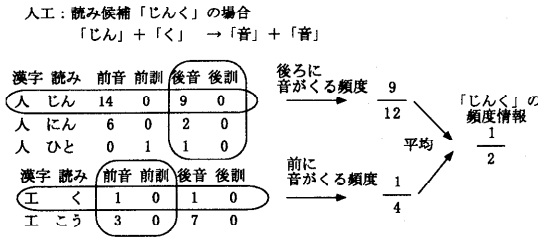


図6 頻度情報の算出例

Fig. 6 The example of calculating frequency.

音訓パターンと読み変化パターンに基づく読みの頻度情報と、品詞の位置関係に基づく意味構成の頻度情報を付加する。各頻度情報は、0~1の範囲で表す。以下に、各頻度の算出方法について述べる。

(1) 音訓パターンに基づく読み頻度情報

n 個の読み (R_1, R_2, \dots, R_n) を持つ漢字 K の読み $R_i \{i: 1, 2, \dots, n\}$ に対し、音または訓の読みが結びつく頻度 $P_{ij} \{j: 0, 1\}$ を、

$$P_{ij} = \frac{r_{ij}}{N} \tag{1}$$

により算出する。 r_{i0} は、読み R_i が音読みと組合せになった回数、 r_{i1} は訓読みと組合せになった回数を表す。 r_{ij} は、漢字知識の「読み知識」を利用し、漢字 K が熟語の前部にある場合は、「後音」と「後訓」、後部にある場合は「前音」と「前訓」を参照して求める。 N は、漢字 K が前部または後部に出現した回数の総和であり、

$$N = \sum_{i=1}^n (r_{i0} + r_{i1}) \tag{2}$$

で求める。なお、出現回数が0の漢字の場合には $N = 0$ になるため、KIDS-IIでは音読みが優先して組み合わせられると仮定し、 $r_{i0} = 1$ とする。

本論文では、式(1)、(2)と読み候補の前部および後部の音訓パターンを考慮して、前部の頻度 $P_{ij}^{(f)}$ と後部の頻度 $P_{ij}^{(b)}$ を算出し、その平均値 F

$$F = \frac{P_{ij}^{(f)} + P_{ij}^{(b)}}{2} \tag{3}$$

を読みの頻度情報とする。図6に、「人工」を対象とした読みの頻度情報の算出例を示す。

(2) 読み変化規則に基づく読み頻度情報

本論文では、読み変化規則に該当する読み候補から、促音、濁音、半濁音をともなった読み候補が導出された場合、この読み候補が元の候補よりも頻度が高いと仮定する。そこで、読み変化をともなった頻度情報 E は、式(3)の F と重み g を用いて、

$$E = F + g \times H \tag{4}$$

と定義する。 g は、変化した読みの頻度を高くする重みであり、 E を、0~1の範囲に制御するために

$$g = \frac{1 - F}{2} \quad (F \geq 0.5) \tag{5}$$

$$g = \frac{F}{2} \quad (F < 0.5) \tag{6}$$

と定義する。 H は、読み変化の発生率であり、該当する読みパターンの出現回数を w 、読み変化が発生した回数を v とすると

$$H = \frac{v}{w} \tag{7}$$

で示される。 v, w は知識ベースの類推規則に該当するテーブルを参照し求める。

(3) 品詞の位置に着目した意味頻度情報

n 個の品詞 (M_1, M_2, \dots, M_n) を持つ漢字 K の品詞 $M_i \{i: 1, 2, \dots, n\}$ が、二字熟語中の前部または後部で利用される頻度 Q_i を、

$$Q_i = \frac{m_i}{N} \tag{8}$$

により算出する。 m_i は、漢字知識の「品詞知識」に収録された出現回数であり、漢字 K が前部にある場合は「前部」、後部にある場合は「後部」のフィールドを参照して求める。 N は、漢字 K が前部または後部で使われた総数で、

$$N = \sum_{i=1}^n m_i \tag{9}$$

で求める。 $N = 0$ の場合は、 $Q_i = 0$ とする。式(8)、(9)から前部と後部の品詞の頻度を算出し、その平均値を意味構成の頻度情報とする。

6. ユーザインタフェース

KIDS-IIの操作環境は、学習者の入力する語が登録語か未登録語かによって、(1) 熟語検索モード、(2) 熟語導出モードに分けられる。(1)では、学習者は正しい知識の他に、検索された熟語に対する類推候補を検索できる。(2)では、学習者は知識ベースに未登録の情報から、知識ベースに存在する熟語を導出できる。以下各モードについて述べる。

6.1 熟語検索モード

熟語検索モードは、(1) 熟語検索ウィンドウ、(2) 関連語検索ウィンドウ、(3) 読み候補導出ウィンドウ、(4) 意味候補導出ウィンドウで構成する。図7に、登録語「人工」における検索例を示す。各ウィンドウと操作環境を以下に述べる。

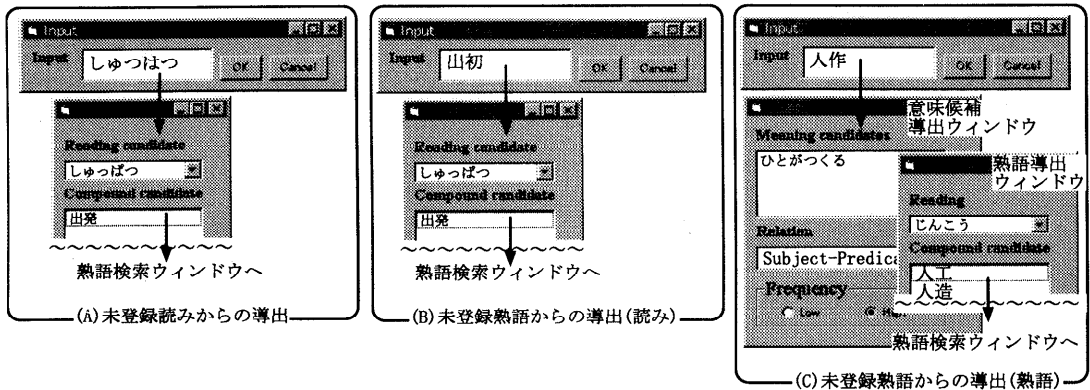


図 8 熟語導出モード例

Fig. 8 The example of compounds derivation mode.

dates フィールド”により構成される。“Relation”は、4.3 節で述べた意味関係名の項目を持ち、システムは、学習者が選択した意味関係名に対応する意味候補を頻度の高い順に表示する。また、表示された意味候補の1つを学習者がマウスで選択し、その意味候補の頻度情報を“Frequency”の“High (頻度>0%)”, “Low (頻度=0%)”で表示し、学習者に自分が類推した意味の妥当性の評価を支援する。

6.2 熟語導出モード

熟語導出モードは、読み候補導出ウィンドウ、意味候補導出ウィンドウ、熟語導出ウィンドウで構成し、(1) 未登録読みからの熟語導出、(2) 読みに着目した未登録熟語からの熟語導出、(3) 意味に着目した未登録熟語からの熟語導出を支援する。図 8 は、各導出の操作例であり、図中の (A) は学習者が「出発」という未習熟語を見て、「しゅっぱつ」という未登録読みを入力した場合、(B) は「しゅっぱつ」という語彙から同音漢字列「出初」を入力した場合である。いずれも、「出発」という正しい熟語を導出している。また、(C) のように学習者が「人作」という未知熟語を類推した場合でも、“主述の関係”に対応する熟語候補「人工」、「人造」という熟語を導出している。従来の電子辞書では、未登録語としていずれも扱えなかったが、KIDS-II はこれらを類推候補として扱い、熟語の導出支援を可能とする。

7. システムの評価と応用

本章では、KIDS-II の類推候補導出機能の実験的評価と、CAI システムへの応用について検証、考察する。

7.1 類推候補導出の評価

我々は、本システムの類推候補導出機能の精度を評価するために、システムが導出した読み候補に対し、

次の実験的評価を行った。

7.1.1 実験内容

(実験 1) 登録熟語に対する読み候補の有効性

類推規則の抽出に利用した知識ベース内の漢字熟語(読み変化規則で表せない熟字訓や連声の読みを持つ熟語を除いた) 2764 語を、(1) 音+音の熟語、(2) 音+訓の熟語、(3) 訓+音の熟語、(4) 訓+訓の熟語、(5) 促音化の熟語、(6) 濁音化の熟語、(7) 半濁音化の熟語の 7 パターンに分け、それぞれに読み導出を行い、正解の読みが候補順位何位に含まれているか評価し、読み規則と頻度情報の有効性を検証した。

(実験 2) 未収録熟語に対する読み候補の有効性

熟語知識に収録されていない 10569 語の漢字熟語を対象に、実験 1 と同じ評価を行い、未収録の熟語に対する読み規則と頻度情報の有効性を検証した。

7.1.2 実験結果と考察

実験 1, 2 の結果を表 4 に示す。表中のパーセンテージは、各パターンに対する正解を含んだ読みの候補順位の割合を示している。

実験 1 では、実験対象すべての熟語に対して正解を導き出し、いずれの読みパターンでも半数以上が候補順位 2 位以上に位置した。これは、学習者が登録語に対して読みの類推を行った場合、その類推が候補順位 3 位以下に位置する候補は、頻度的に希な類推であると評価できる。

実験 2 では、実験 1 と比べ全体的に精度は低下しているが、候補順位 4 位以下に正解が存在する割合はいずれのパターンでも 30% 以下であった。類推規則抽出に用いたデータが 2832 語と少ない点を考慮すれば、未収録語に対しても良好な結果を得たと考える。これは、熟語導出モードで登録熟語の導出に失敗した場合でも、学習者は自らの読みの類推に対して登録熟語と

表4 実験結果
Table 4 Results of experimentation.

読みパターン	候補順位	実験1	実験2
音+音	1位	89%	88%
	2位	9%	9%
	3位	1%	2%
	4位以下	1%	1%
音+訓	1位	25%	18%
	2位	33%	18%
	3位	25%	33%
	4位以下	17%	31%
訓+音	1位	56%	18%
	2位	16%	23%
	3位	19%	30%
	4位以下	9%	29%
訓+訓	1位	44%	8%
	2位	26%	26%
	3位	19%	51%
	4位以下	11%	15%
促音	1位	96%	94%
	2位	1%	2%
	3位	2%	3%
	4位以下	1%	1%
濁音	1位	70%	53%
	2位	10%	16%
	3位	18%	23%
	4位以下	2%	8%
半濁音	1位	100%	96%
	2位	0%	2%
	3位	0%	1%
	4位以下	0%	1%

同様に類推を評価できることを示している。なお、読み候補が導出できなかった語が促音化の熟語に572語中103語、濁音化の熟語に402語中306語存在した。これらは、KIDS-IIの持つ読み変化規則の未対応によるものであるが、読み変化規則を追加していくことでこれらの熟語に対応できる。

以上の実験結果は読みの類推学習支援におけるKIDS-IIの算出する頻度情報の有効性を示していると考えられる。

なお、意味候補の有効性については、漢字の構成と意味に着目して熟語の意味を捉えた場合、その意味構成が一意に決まらない語や、派生する意味を持つ語が多く、正解との比較に基づく厳密な評価が期待できないため、評価実験を行わなかった。

今後の実験課題として、外国人留学生を対象とした本システムの試用評価があげられるが、試用を通じて、意味候補の有効性や学習効果について検証したい。また、今回の評価は類推規則抽出に用いたデータが2834語と少なく、実験的なものであるため、データの追加、および洗練化による導出精度の向上についても評価を行う必要がある。

7.2 CAIシステムへの応用

KIDS-IIは、漢字熟語の類推を支援する学習用辞書システムとしてだけでなく、CAIシステムへの応用として、(i) 知識ベースとしての利用、(ii) 学習者モデルとしての利用が考えられる。

現在、我々はKIDS-IIをベースにして、ゲーム型漢字熟語CAI“JUGAME”¹⁴⁾の改良版であるJUGAME-II構築を行っている。JUGAME-IIは、学習者主体の“ゲームモード”の他に、学習者に問題を提示する“演習モード”が用意されている。各モードでのKIDS-IIの利用方法について以下に述べる。

“ゲームモード”は、JUGAMEの学習形態を継承した学習環境である。JUGAMEは、学習者が知識ベースに存在しない熟語を作成したとき、「辞書に登録されていない」という応答しかできなかった。そのため、学習者が自らの作成した熟語の正当性が評価できない問題があった。しかしJUGAME-IIでは、KIDS-IIを知識ベースに利用し、学習者に作成した熟語の読み、意味を入力させるタスクを与え、類推候補と比較し、未登録語についても学習者の自分の回答に対する評価が可能になった。

“演習モード”は、ゲームモードに熟語の穴埋め形式の演習問題を加えた学習環境である。知的教育システムにおける学習者モデルの構築手法の1つにパーテーション法がある。学習者の誤りを正しい知識の変形と考え、変形の形態をオペレータとして整理し、正しい知識に適用して学習者の振舞いを説明する枠組みである¹⁵⁾。KIDS-IIが導出する読みと意味の類推候補を、類推のオペレータによる導出結果と捉えれば、問題設定がされている状況下では、実際の読み以外の読み候補は、誤りのオペレータによるバグといえる。この考えに基づき、KIDS-IIを学習者モデルの構築に利用したことにより、演習モードは、単なるドリル&プラクティスの枠組みではなく、学習者の誤りの診断・分析が可能となった。JUGAME-IIの詳細については、別の機会に論じる。このように、KIDS-IIの応用形態は単なるCAIの知識ベースの枠組みだけでなく、学習者モデルへの応用も可能である。

8. むすび

本論文は類推候補を扱う漢字熟語の知識ベースの提案と、これを用いて試作した漢字熟語辞書システム“KIDS-II”について述べた。熟語の読みや意味を単漢字の知識と関連づけて理解することは教育的にも重要である。KIDS-IIは、二字熟語2832語、常用漢字1945字を扱い、熟語の意味と読みの類推規則を用い

て、学習者の類推に対応する意味候補、読み候補、熟語候補の導出と、類推結果の評価情報の提示が可能である。本システムは、学習者の類推を支援する漢字熟語辞書としての利用が期待できる。また、CAIシステムへの応用についても、知識ベースとしてだけでなく、学習者モデルへの応用も可能である。KIDS-IIはパーソナルコンピュータ上に試作した。OSはMicrosoft Windows95、開発言語はVisual Basicを用い、データベースの作成にはMicrosoft Accessを用いた。

謝辞 本論文の執筆時に貴重なご意見をいただいた福井大学教育学部脇田里子講師に深謝します。なお、本研究の一部は、平成9年度科学研究費補助金基盤研究(B)(2)展開研究(No.09558017)の補助を受けている。

参考文献

- 1) 大坪一夫：日本語教育でのコンピュータ利用の過去、現在と未来、日本語教育、No.78, pp.9-19 (1992).
- 2) 森田良行：日本語学と日本語教育，凡人社(1990).
- 3) 石田敏子：日本語教授法，大修館書店(1992).
- 4) 木村宗男：読解の指導方法，日本語教授法の諸問題，日本語教育指導参考書3，pp.1-59，文化庁(1972).
- 5) 川口義一，加納千恵子，酒井順子：日本語教師のための漢字指導アイデアブック，創拓社(1995).
- 6) 藤堂明保：漢語と日本語，秀英出版(1969).
- 7) 波多野諠余夫，小島恵子：未知複合語の意味推定と心内辞書登録，電子情報通信学会第二種研究会資料，LK92-10，pp.1-8(1992).
- 8) 浅木森利昭：マルチメディアを利用した日本語教育支援システムの開発，科学研究成果報告書(試験研究(B)(1))(1994).
- 9) 林 敏浩，矢野米雄，三好克美，関 康夫：漢字熟語学習支援のための電子辞書システムの試作，CAI学会論文誌，Vol.10, No.4, pp.159-170(1994).
- 10) 亀田弘之，波多野諠余夫，小島恵子：人間的な類推に基づく未知語の意味推定システム，電子情報通信学会技術研究報告「思考と言語研究会」資料，TL95-2，pp.9-20(1995).
- 11) Ochi, Y., Yano, Y. and Hayashi, T.: The development of a Kanji compound dictionary system for deriving their meaning and reading, *International Conference on Computers in Education ICCE'95, Singapore*, pp.549-556 (1995).
- 12) 吉村賢治，首藤公昭：2漢字語の解析，福岡大学工学集報，第49号，pp.229-237(1992).
- 13) 天沼 寧，大坪一夫，水谷 修：日本語音声学，くろしお出版(1978).
- 14) 矢野米雄，林 敏浩，三好克美，関 康夫：ゲームスタイルを用いた漢字熟語学習システム，CAI学会誌，Vol.11, No.2, pp.75-85(1994).
- 15) 溝口理一郎：知的教育システム，情報処理学会誌，Vol.36, No.2, pp.177-186(1995).

(平成8年11月29日受付)

(平成9年11月5日採録)



越智 洋司(学生会員)

1971年生。1994年徳島大学工学部知能情報工学科卒業。1996年同大学院工学研究科知能情報工学専攻博士前期課程修了。現在、同大学院博士後期課程在学中。日本語教育支援システムの研究開発に従事。教育システム情報学会会員。



矢野 米雄(正会員)

1945年生。1969年大阪大学工学部通信工学科卒業。1974年同大学院工学研究科博士課程修了。工学博士。同年徳島大学助手。1975年同講師。1981年同助教授。1990年同教授。1979～1980年米国イリノイ大学 Computer-based Education Research Laboratory 客員研究員。環境型知的CAIシステム，人脈活用支援システム，ゲーム環境の研究に従事。教育システム情報学会副会長・理事・編集委員長・CAI研究部会長。日本教育工学協会理事。電子情報通信学会，米国IEEE各会員。



林 敏浩(正会員)

1966年生。1989年徳島大学工学部情報工学科卒業。1994年同大学院工学研究科博士後期課程修了。博士(工学)。同年佐賀大学理工学部講師。1996年同助教授。環境型教育システム，マルチメディア教育システム，ネットワーク型教育システムなどの研究に従事。教育システム情報学会，電子情報通信学会，日本教育工学会，人工知能学会，米国AAACE各会員。