

平成 26 年 5 月 29 日現在

機関番号：35404

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23530966

研究課題名(和文) 漢字二字熟語間の意味的関連性に関するデータベースの構築

研究課題名(英文) Database of semantic transparency and semantic relatedness of Japanese two-kanji compound words

研究代表者

増田 尚史 (MASUDA, Hisashi)

広島修道大学・人文学部・教授

研究者番号：90335018

交付決定額(研究期間全体)：(直接経費) 3,900,000円、(間接経費) 1,170,000円

研究成果の概要(和文)：漢字二字熟語10,015語を対象に、熟語とそれを構成している左右それぞれの漢字間の意味的関連性(意味的透明性)を大学生に評定させた。その結果、中程度以上の意味的透明性を持つ語が全体の94%を占めることが確認された。また、熟語の統語構造や音訓構造、さらには漢字が表示している形態素の違いが意味的透明性に影響を与えることが示された。意味的透明性のデータに基づき、同一の漢字を共有する熟語群間の意味的関連性を規定し、これらをデータベース化した。

研究成果の概要(英文)：In the present project, semantic transparency for Japanese 10,015 two-kanji compound words was estimated by undergraduate students. The results showed that 94% of words are transparent (i. e., semantically related to both kanji characters constructing the words). The results also indicated that syntactic structure and phonological structure, as well as morphology each kanji denotes, affect semantic transparency. Based on the semantic transparency data, semantic relatedness among words those have a same kanji character as a component was calculated. Finally, database of the semantic transparency and the semantic relatedness was constructed.

研究分野：社会科学

科研費の分科・細目：心理学・実験心理学

キーワード：漢字二字熟語 意味的関連性 意味的透明性 データベース

## 1. 研究開始当初の背景

(1) 読み (reading) の活動を支えている単語認知過程を心理学的に検討する上では、実験刺激として用いる言語材料の精選が不可欠である。そのためには、単語が持っている種々の属性を抽出し、データベース化しておく必要がある。わが国においては現在までに、国立国語研究所を中心にして、幾種類かのデータベースあるいは言語コーパスが作成されている。しかしながら、その多くは単語の出現頻度や用例をデータベース化したものであり (たとえば、国立国語研究所の『現代雑誌 200 万字言語調査彙表』や NTT データベースシリーズ『「日本語の語彙特性」第 II 期』)、個別の単語に対する心理量を測定したものは極めて少ない。例外として、NTT データベースシリーズ『「日本語の語彙特性」第 I 期』および『第 III 期』があり、それぞれ単語の親密度と心像性をデータベース化しているが、単語の“意味”を直接的に扱ったものではない。

(2) 近年、日本語以外の諸言語を実験材料とする単語認知研究において、意味的透明性 (複合語の意味とその語を構成している単純語の意味との関連性) が、単語の認知過程に大きく影響することが報告されている。このことは、より厳密には、一つの単語の意味だけではなく、複数の単語間の意味的関連性が単語認知過程に関与することを示している。したがって、日本語の単語についても、その形態情報や音韻情報とともに、意味情報に関するデータベースが必要不可欠となってきた。

(3) 日本語の単語についても、既に、たとえば『日本語語彙大系』や『EDR 電子化辞書』においてデータベース化の試みはなされているが、これらは、アプリアリに決められた意味カテゴリのいずれに、各単語が属するかを表現したシソーラスであり、人間の心理量は測定されていない。

## 2. 研究の目的

『広辞苑』に見出し語として記載されている、日本工業規格第一水準の漢字 (2,965 字) から構成されている漢字二字熟語 (78,000 語以上) のうち、出現頻度の相対的に高い 10,000 語余りに絞って、以下のデータベースを構築する。

### (1) 意味的透明性 (semantic transparency)

熟語の意味と、それを構成している漢字の意味との関連性の程度を調べ、データベース化する。

### (2) 意味的関連性 (semantic relatedness)

意味的透明性のデータを元に、同一の漢字を含む熟語間の意味的関連性を規定し、データベース化する。

## 3. 研究の方法

### (1) 調査協力者

研究代表者および分担者の所属機関の大学生延べ 1710 名が調査に参加した。

### (2) 調査対象とした漢字二字熟語

常用漢字 (2,136 字) のうち、日本工業規格第一水準に登録されている 2,106 字の組合せからなる二字熟語から、国立国語研究所の現代日本語書き言葉均衡コーパスを基に算出した長単位出現頻度 (Joyce, Hodošček, & Nishina, 2010, 2012) が 49 以上の 10,015 語を、調査対象熟語として選定した。

### (3) 調査用熟語リストの作成

選定した 10,015 語を、100 語もしくは 101 語からなる 100 種類のリストに分割し、各リストを調査の最小単位とした。分割に際しては、各リストに出現頻度の高い語も低い語も含まれ、また平均出現頻度数がリスト間ではほぼ均衡するように配慮した。

### (4) 手続き

①調査はインターネット・ウェブを経由して実施され、調査協力者は 1 リストを単位として回答し、回答結果は自動的にサーバに保存された。②調査協力者には、回答に先立ち、調査の概要に関する説明文を読ませ、調査協力への同意を得た。また、性別と年齢を入力させた。③調査協力者の課題は、まず、画面に提示される熟語の読み方と意味のそれぞれの知識の有無について、2 件法 (知っている/知らない) で回答することであった。次に、読みと意味のいずれについても“知っている”と回答した場合にのみ、熟語の意味と、その第 1 文字目 (左漢字) および第 2 文字目 (右漢字) の意味との関連性について、それぞれ 6 件法 (非常に/かなり/少し関連性がある、あまり/ほとんど/全く関連性がない) で回答することであった (図 1 を参照)。④リスト内の単語の呈示順序は協力者ごとに無作為化された。1 リストあたりの回答作業時間は、いずれの協力者も 30 分以内であった。⑤各調査協力者は最低でも 1 種類のリストのすべての熟語について回答した。また、いずれの熟語に対しても、最低でも 13 名の協力者から回答を得た。

過重

q1 この熟語の読み方を

知っている 知らない

q2 この熟語の意味を

知っている 知らない

q3 熟語の意味と左側の漢字の意味とは

非常に関連性がある かなり関連性がある 少し関連性がある  
あまり関連性がない ほとんど関連性がない 全く関連性がない

q4 熟語の意味と右側の漢字の意味とは

非常に関連性がある かなり関連性がある 少し関連性がある  
あまり関連性がない ほとんど関連性がない 全く関連性がない

次へ

図1 回答用ウェブページ

4. 研究成果

(1) 既知率と親密度・出現頻度

熟語の読み方および意味の既知率と、熟語の親密度(天野・近藤, 1999)、出現頻度(天野・近藤, 2000; Joyce, Hodošček, & Nishina, 2012)の対数変換値との相関係数は、表1の通りであった。熟語の読み方および意味の既知率が、親密度との相関は低かったものの、出現頻度の対数変換値とは高い正の相関関係を示したことは、本調査の協力者が無秩序に回答していたわけではないことを意味する。

表1

熟語の属性	既知率	
	読み方	意味
親密度(天野・近藤, 1999)	.05	.06
出現頻度		
天野・近藤(2000)	.37	.40
Joyce et al. (2012)	.32	.39

(2) 熟語の意味的透明性の分布

熟語の意味と、その構成要素である漢字の意味との関連性(意味的透明性)に関する評定について、“全く関連性がない”を0点とし、“非常に関連性がある”を5点として得点化した。以下では、熟語ごとに得られた平均評定値を、意味的透明度と呼ぶ。なお、2種類の熟語(“群落”、“馬丁”)については、回答したすべての協力者(それぞれ17名)が読み方もしくは意味を“知らない”と回答したため、以下の分析からは除外された。表2は、左右の構成漢字に対する意味的透明度の区間ごとの熟語数を示す。また、表3は左右それぞれの漢字に対する意味的透明度の高低で2分割した際の該当する熟語数と、それぞれの意味的透明度の平均値を示す。表2および表3から、左右いずれの漢字に対する意味的透明度も4以上の熟語が、全体の62%以上を占め、同様に3以上の熟語が全体の94%

以上を占めることが確認された。しかしながらその一方で、少数ながらも、構成漢字に対する意味的透明度が低い熟語も存在することが確認された。たとえば、左漢字に対する意味的透明度は高いが、右漢字に対するそれが低い熟語には“名前、克服”などがあり、逆に、左漢字に対する意味的透明度は低いながら、右漢字に対するそれは高い熟語には“喫緊、活字”などがあつた。さらに、左右いずれの漢字に対する意味的透明度も低い熟語には、“科白、人参”などがあつた。

さらに、意味的透明度の分散に注目すると、たとえば“右翼、綱領”などの熟語では、その左漢字に対する意味的透明度の分散が相対的に大きく、また“欧米、相棒”などの熟語では、その右漢字に対する意味的透明度の分散が相対的に大きかった。これらは、漢字の多義性(たとえば、“米”=こめ、アメリカ)や、熟語および構成漢字に対する評定者の知識のばらつきを反映していると考えられる。

表2  
左右の構成漢字との意味的透明度の区間ごとの熟語数

左漢字	右漢字				
	5~4	4~3	3~2	2~1	1~0
5~4	6,213	1,583	179	2	0
4~3	1,082	577	102	6	0
3~2	107	88	49	5	0
2~1	2	3	5	9	0
1~0	0	0	0	0	1

表3  
左右構成漢字との意味的透明度の高低による熟語数と各意味的透明度

左漢字	右漢字			
	5~3		3~0	
5~3	N = 9455 (94.4%)		N = 289 (2.9%)	
	左 = 4.4	右 = 4.3	左 = 4.1	右 = 2.6
3~0	N = 200 (2.0%)		N = 69 (0.7%)	
	左 = 2.6	右 = 4.0	左 = 2.3	右 = 2.3

(3) 統語構造と意味的透明度

漢字二字熟語の統語構造と意味的透明度との関連性を検討するために、Masuda & Joyce (2005)のデータベースにおいて、熟語の統語構造(MM = 修飾語+被修飾語、CV = 補語+動詞、VC = 動詞+補語、SP = 同義語反復)が明確である語と重なっている4,918語について、統語構造ごとに意味的透明度を求めた(表4を参照)。その結果、同義語反復構造を有する熟語(たとえば、“恋愛、平等”)の意味的透明度が、左右いずれの漢字に対しても高いことが確認された。このことは、同義語反復が、それぞれの漢字に対する意味的透明度に対して相乗効果をもたらす

ことを示す。

表4  
熟語の統合構造ごとの意味的透明度の平均値(標準偏差)と  
左右漢字に対する意味的透明度の大小関係ごとの熟語数

統語構造	意味的透明度		意味的透明度の大小関係 ごとの熟語数		
	左漢字	右漢字	左>右	左=右	左<右
	MM (N = 3548)	4.4 (0.47)	4.3 (0.49)	1,870	225
CV (N = 686)	4.3 (0.45)	4.3 (0.43)	329	37	320
VC (N = 601)	4.4 (0.39)	4.3 (0.51)	307	46	248
SP (N = 83)	4.7 (0.24)	4.6 (0.31)	48	10	25

#### (4) 音訓構造と意味的透明度

漢字の形態素としての役割に注目し、読みが異なれば形態素が異なると仮定し、熟語の音訓構造(音音=音読み+音読み、音訓=音読み+訓読み、訓音=訓読み+音読み、訓訓=訓読み+訓読み)ごとに意味的透明度を求めた(表5を参照)。その結果、音訓構造(いわゆる、重箱読み)と訓音構造(いわゆる湯桶読み)の熟語においては、左右いずれの漢字に対しても意味的透明度が相対的に低いことを確認した。このことは、和語に由来する形態素と中国語に由来する形態素とが結合した熟語では、意味的透明性が低い傾向にあることを意味する。

表5  
熟語の統合構造ごとの意味的透明度の  
平均値(標準偏差)

音訓構造	意味的透明度	
	左漢字	右漢字
音音 (N = 8690)	4.3 (0.50)	4.2 (0.53)
音訓 (N = 130)	4.1 (0.65)	4.0 (0.74)
訓音 (N = 157)	4.3 (0.53)	4.0 (0.64)
訓訓 (N = 930)	4.3 (0.56)	4.2 (0.60)

#### (5) 形態素と意味的透明度に関する事例分析

上記(4)と同様に、読みが異なれば形態素が異なると仮定すると、たとえば、2種類の音読みを持つ漢字については、熟語内においてそのいずれの読み方がなされているかによって、表示している形態素が異なると考えられる。表6は、この基準にしたがって形態素を同定した際の、熟語内の左右それぞれの位置を占める形態素数を、音読みと訓読み別に示す。ただし、算出に当たっては、熟字訓のように漢字が表示している形態素が不明確な熟語を除外した。表6から、複数の熟語を構成している形態素が、左右いずれも全体の約66%を占めていることがわかる。

表6  
熟語内の左右位置を占める形態素の数

読み	左形態素		右形態素	
	熟語数 > 1	熟語数 = 1	熟語数 > 1	熟語数 = 1
音読み	1180	496	1113	502
訓読み	209	201	193	161

たとえば、漢字“下”が語頭(左漢字)に位置して/ka/と音読みされる熟語(N = 15; “下位、下降”など)の意味的透明度の平均値は5.5(SD = .25)であり、/ge/と音読みされる熟語(N = 17; “下山、下車”など)の意味的透明度の平均値は4.9(SD = .42)であった。また、/jita/と訓読みされる熟語(N = 12; “下請、下味”など)の意味的透明度の平均値は4.8(SD = .38)であった。これらのことは、熟語を構成する漢字が表示している形態素の違いによって、熟語の意味的透明性が異なることを示唆する。

さらに、同じ“下”が語尾(右漢字)に位置する熟語では、/ka/ (N = 25; “以下、部下”など)、/ge/ (N = 1; “上下”)、および/jita/ (N = 6; “靴下、年下”など)と発音される熟語の意味的透明度の平均値は、それぞれ5.3(SD = .41)、5.6(SD = .00)、5.2(SD = .60)であった。このことは、同一の形態素を表示している漢字であっても、熟語内に占める位置によって、熟語の意味的透明性に影響を与えることを示唆する。

#### (6) 熟語間の意味的関連性

意味的透明度は、熟語とそれを構成している漢字との意味的関連性を表すことから、同一の漢字を構成要素とする2種類の熟語の当該漢字に対する意味的透明度の差を、熟語間の意味的距離と見なす。そこで、この方式によって、熟語間の意味的距離を求めた上で、これを計算上の最大値である5から差し引いた値を、当該熟語間の意味的関連度と見なした。たとえば、“楽天”の左漢字“楽”に対する意味的透明度は4.5であり、“楽園”のそれは4.2であることから、両熟語間の意味的距離は0.3であり、意味的関連度は4.7と計算される。さらに、“楽”を右漢字とする“道楽”の意味的関連度は4.0であることから、“楽園”と“道楽”の間の意味的距離を0.2、意味的関連度を4.8と見なすことができる。ただし、上記(5)の分析結果に基づき、“楽園”や“道楽”のように“楽”を/raku/と発音する熟語と、“楽曲”や“音楽”のように“楽”を/gaku/と発音する熟語とは、異なる形態素からなると想定されるので、別々に意味的関連度を計算した。

次に、同じ形態素を熟語内の左右同じ位置に含む熟語群において、意味的関連度の最小値を求め、これを当該熟語群の意味的関連度と見なした。たとえば、漢字“文”が語頭(左漢字)に位置して/bun/と音読みされる熟語群(N = 23)のうち、“文体”の意味的透明度が最も高く(4.9)、“文壇”のそれが最も低い(3.4)ので、これらの差(1.4 = 4.9 - 3.4)を5から差し引いた値(3.6 = 5 - 1.4)を、この熟語群の最小意味的関連度と見なした。表7は、同一形態素を共有する熟語群の最小

意味的関連度を示す。

表 7

同一形態素を共有する熟語群の最小意味的  
関連度の平均値 (標準偏差)

読み	左形態素	右形態素
音読み	4.16 (0.65)	4.00 (0.64)
訓読み	4.31 (0.68)	4.24 (0.69)

(7)以上の分析から得られた、意味的透明度と、同一形態素からなる熟語群での最小意味的関連度を、出現頻度や統語構造などその他の指標とともにデータベース化した。現在、ホームページ ([https://www.researchgate.net/profile/Hisashi\\_Masuda](https://www.researchgate.net/profile/Hisashi_Masuda)) での公開に向けて準備している。

#### (8)今後の展望

漢字二字熟語は複合語であることから、その形態素構造や統語構造が、語の認知に影響を与えていることが予想される。しかしながら、これまでの日本語の視覚的単語認知研究においては、漢字熟語を複合語と捉える観点が欠如しており、研究例はきわめて稀であった。例外的な研究例として、Joyce & Masuda (2008, 2009)は統語構造が語の認知に及ぼす影響を検討し、“動詞+補語”や“補語+動詞”の構造を有する熟語では、主要部 (head) である動詞の役割を持つ漢字を先行呈示することにより、相対的に迅速に熟語全体の認知が遂行されることを確認している。しかしながら、“修飾語+被修飾語”の構造を有する熟語では、主要部である被修飾語に相当する右漢字の先行呈示が必ずしも語の認知を促進するとは限らなかった。しかし、本研究課題によって得た意味的透明度の分布によれば、“修飾語+被修飾語”の構造を有する熟語では、修飾語に相当する左漢字に対する意味的透明度の平均値が右漢字に対するよりも高く、またそのような熟語数が多いことが確認された (表 4 を参照)。このことが Joyce & Masuda (2008, 2009)の実験結果に影響を与えたと考えられる (Joyce & Masuda, 2013)。

このように、本研究課題で得られたデータに基づいて、統語構造に加えて意味的透明性や熟語間の意味的関連性の観点からも熟語を選定して、認知実験を実施すれば、熟語がいかなるプロセスで認知されているのか、より詳細に検討できると予想される。さらに、英語などと比較して、日本語には漢字二字熟語に代表されるように非常に多くの複合語が存在するので、本研究課題によるデータベースに基づいて選定された熟語を用いた研究によって、他の言語では検討しえない認知プロセスが明らかになることが期待される。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 0 件)

[学会発表] (計 8 件)

① 増田尚史 複合語としての漢字熟語の意味と認知 第 17 回認知神経心理学研究会 2014 年 8 月 23 日 岡山県立図書館多目的ホール

② 増田尚史・藤田知加子・小河妙子・Terry Joyce・川上正浩 漢字二字熟語の意味的透明性に関する調査 日本基礎心理学会第 32 回大会 2013 年 12 月 8 日 金沢市文化ホール

③ Joyce, T. & Masuda, H. Constituent-morpheme priming study of Sino-Japanese and Native-Japanese compound words. 8th International Morphological Processing Conference 2013 年 6 月 22 日 Cambridge, UK.

④ 増田尚史・Terry Joyce・小河妙子・藤田知加子・川上正浩 メンタル・レキシコン研究 (XVI) : 漢字二字熟語の意味的透明性のデータベース構築に向けて 日本心理学会第 76 回大会 2012 年 9 月 13 日 専修大学

[図書] (計 0 件)

[産業財産権]

○出願状況 (計 0 件)

○取得状況 (計 0 件)

[その他]

ホームページ等

#### 6. 研究組織

##### (1) 研究代表者

増田 尚史 (MASUDA, Hisashi)  
広島修道大学・人文学部・教授  
研究者番号：90335018

##### (2) 研究分担者

ジョイス、テリー (JOYCE, Terry)  
多摩大学・グローバルスタディーズ学部・教授  
研究者番号：20418677

小河 妙子 (OGAWA, Taeko)  
東海学院大学・人間関係学部・准教授

研究者番号： 3 0 4 3 4 5 1 7

川上 正浩 (KAWAKAMI, Masahiro)  
大阪樟蔭女子大学・心理学部・教授  
研究者番号： 4 0 2 4 2 7 8 9

藤田 知加子 (FUJITA, Chikako)  
南山大学・人文学部・准教授  
研究者番号： 7 0 3 0 0 1 8 4