

# 韓国語の 2 字漢字語の語彙性アスペクトによる動詞化および形容詞化の予測

朴善嫻（名古屋大学大学院）・玉岡賀津雄（名古屋大学）・李在鎬（筑波大学）

## 要旨

韓国語の均衡コーパスから抽出した漢字 2 字で構成される上位 2,000 語の高頻度語を対象に、動詞化または形容詞化を予測する二項ロジスティック回帰分析を行った。動詞化・形容詞化の可否は辞書に準じて行った。予測に使用した変数は、4 つの語彙性アスペクト（「開始」「継続」「終結」「状態」とコーパスでの出現頻度である。分析の結果、予測力は、動詞化には「終結」が強めて強く、それに「状態」、形容詞化には「状態」が強かった。

## 1. はじめに

本研究では、韓国語の「식사（食事）/sigsa」「운동（運動）/undong」のように 2 つの漢字で構成された「2 字漢字語」を対象に調査を行う<sup>1</sup>。調査に先立ち、韓国語における語彙の分布および日本語との関連について述べる。

韓国語の語彙は、固有語、漢字語、外来語に分類できる。約 40 万語の現代韓国語を収録している『国語大辞典』の語彙比率では、固有語が 25.90%、漢字語が 67.02%<sup>2</sup>、外来語は 7.08%である（李, 1980）。さらに、辞書に収録されていない学術用語を含めると漢字語の比率はさらに高くなるであろう。

韓国語と日本語では、漢字 2 字からなる語彙に関して、共通する語彙が多く、いずれの言語においても頻繁に使われている。日本語の場合、国語辞典に記載された見出し語の内、約 70%が漢字 2 字で構成されている（Yokosawa & Umeda, 1988）。ただし、これらが全て漢語というわけではないが、日本語では漢字表記の語彙が多く使われていることが分かる。一方、韓国語の場合、ハングルによって音韻表記されるため、日常生活で漢字を目にすることはほとんどない。唯一の例外は、意味的区別が求められる同音意義語に対して使用される場合である。例えば、「이상/isang/」の発音に相当する語は、全部で 12 語あり、高頻度語としては「以上/isang/」「理想/isang/」「異常/isang/」などが挙げられる。前後の文脈だけで区別が難しい場合、「…이상（理想）…」のように漢字を併記することがあることから、ハングルで表記されていても、それが漢字に置き換えて表記できることは母語話者によって意識されている。こうした事実から、表記の問題とは別に、2 字漢字語を認定することができると考える。

<sup>1</sup> 本稿における韓国語の音素表記は、〈文化観光部告示第 2000-8 号（2000.7.7）〉に従って行う。

<sup>2</sup> この比率には、〔固有語＋漢字語〕や〔漢字語＋外来語〕も含まれている。〔固有語＋漢字語〕や〔漢字語＋外来語〕と漢字のみで構成されている語を区別するために、本研究では、漢字のみで構成されている語を「漢字語」と呼ぶ。

韓国語と日本語は、ともに膠着語であること、用言類の形態的屈折によって文法関係を表すことから、漢字語に関する文法形式も多分に類似している。例えば、日本語では、漢字表記の名詞に、*-suru* (サ行変格活用動詞、以下、サ変動詞) を付加することで、動詞として使うことができる。韓国語においても同じ現象が存在する。韓国語の場合、*-hada* を付加することで動詞として使用することができるのである。韓国語と日本語におけるこうした類似性から、「体操」「安心」「見学」など、韓国語の「漢字語+*hada*」の形式で動詞化できる語は、日本語でも「漢字語+*suru*」で動詞化できる場合が多い。

さて、日本語の場合、漢語が *-suru* 付加によって動詞化されるかどうかは、その語の持つ語彙性アスペクトが影響していると言われている。具体的な研究として松岡 (2004) では、「開始」「継続」「終結」「状態」の4つの語彙性アスペクトにより、漢字2字で表記される語彙の動詞化がほぼ予測できるという分析結果を明らかにしている。実際、Tamaoka, Matsuoka, Sakai, & Makioka (2005) は、1985年から1998年の間に出版された14年分の朝日新聞のデータから、使用頻度の高い2,000の漢字2字熟語を選んで動詞化する語を調査した。調査の結果、動詞化するのは802語で、全体の40.10%であることが明らかになった。さらに、二項ロジスティック回帰分析を使って、2,000語の動詞化を4つの語彙性アスペクトによって予測した。その結果、4つすべてのアスペクトが *-suru* 付加による動詞化を有意に予測した。しかし、「終結」のアスペクトを持つのは、調査対象の2,000語の内759語で、さらに759語の内751語が動詞化した。つまり、8語だけ(「葬儀」「犯行」「軍縮」など)が、*-suru* 付加による動詞化が成立しなかった。2,000語中、802語の動詞化する語彙の予測率は93.64% (751/802) と極めて高かった。予測力の指標であるWald統計量も312.32であり、非常に強かった。もちろん他のアスペクトも有意な予測変数であったが、上記の予測率からも分かるように、動詞化する漢字2字の語彙のほとんどは「終結」のアスペクトを持っていた。そのため、「終結」単独での予測率が、他のアスペクトと比べて高くなった。

それでは、言語類型的に日本語と類似した特徴を持つ韓国語の2字漢字語でも語彙性アスペクトで、動詞化を予測できるのであろうか。ここで一つ注意しておかななくてはならないことがある。日本語の場合、用言の末尾が *-u* か *-i* かで、動詞であるか形容詞であるかが判断できるが、韓国語の場合、このような形態的な区別は存在しない。本研究の調査対象である漢字語の動詞化と形容詞化においてもこの問題が関連する。というのは、動詞化・形容詞化のいずれにおいても、原形(日本語では、終止形または辞書形と呼ばれている)が *-hada* になるからである。両者を区別する方法は2つ存在する。第1の方法として、活用形によって区別させることができる。例えば、現在時制の連体形の場合、動詞は *-haneun* となり、形容詞は *-han* となるので、両者を区別することができる<sup>3</sup>。

<sup>3</sup>日本語の「合格する方法」は、韓国語では「합격하는 방법 (habgyeog-*haneun*

第2の方法として、意味で区別することができる。形容詞の場合、属性や恒常的性質を表すことが多く、動詞の場合、動作に関連する事象を表すことが多いため、ある程度区別することができるものの、例外もかなり見られ、厳密な区別ではない。

以上を踏まえ、本研究では、動詞化と形容詞化を別々に検討することにした。具体的には、次の手順で研究を進めた。第1に、韓国の『21世紀世宗計画』のコーパスから使用頻度の高い順に2,000語を選んだ。第2に、韓国語の『標準韓国語大辞典』を使って、これらの語彙に *-hada* が付加されるかどうかを記録した。さらに、確認のため、韓国語母語話者の主観判定によって、辞書の記述がどのくらい母語話者の直感と合っているかについても検討した。第3に、これら2,000語についての「開始」「継続」「終結」「状態」の語彙性アスペクトの有無を複数の韓国語母語話者の内省で判定し、判定者間信頼性も検討した。第4に、二項ロジスティック回帰を使って、4種類のアスペクトと使用頻度が、*-hada* 付加による動詞化と形容詞化をどのくらい予測するかを検討した。

## 2. 語彙性アスペクト

動詞をアスペクトに基づいて分類することで、動詞の性質が説明できる。Vendler (1967) は、アスペクトの観点から英語の動詞が表す内的時間の性質を図式化するために、現在進行形の *be+~ing* と共起するか否かと完了点を内包するの否かを分類指標にして、4つの類型を提案した(日本語の説明は、平嶋(2003))。それらは、「達成 (accomplishment) 動詞」「活動 (activity) 動詞」「到達 (achievement) 動詞」「状態 (state) 動詞」である。また、日本語では、金田一(1976) がアスペクトの観点から動詞を4種類に分けている。状態を表わす「状態動詞」、ある時間を継続して行われる種類の動作・作用を表す「継続動詞」、瞬間に終わってしまう動作・作用を表す「瞬間動詞」、時間の概念を含まずある状態を帯びることを表し、常に「~ている」の形である「第四種の動詞」である。

さらに、油谷(1978) は、韓国語の動詞をアスペクトで分類している。まず動詞は「状態性」「結果性」「瞬間性」を持っていると判断し、それを「±状態性」「±結果性」「±瞬間性」のアスペクト素性を持っているか否かで、A類「+状態、+結果」、B類「+状態、-結果」、C類「-状態、+瞬間、+結果」、D類「-状態、+瞬間、-結果」、E類「-状態、-瞬間、+結果」、F類「-状態、-瞬間、-結果」の6つに分類している。このようにアスペクトの観点から動詞を分類すると、複数のアスペクトを共有している場合もあれば、F類のように「状態・瞬間・結果」のどのアスペクトにも当てはまらない場合もある。

英語、日本語、韓国語について、語彙性アスペクトとして共通すると思われ

---

bangbeob)」、形容詞の場合は「正直한人」が、「정직한 한 사람 (jeongjig-han salam)」となる。

るのは、動作性のアスペクトとして活動性、達成性、継続性、瞬間性であり、これに状態性を加えた5つであろう。語彙性アスペクトをどう呼ぶか、またどのくらいの粒度で分類するかは、個々の研究によって多少の違いがあるものの、基本的なアプローチとして、ある動作が始まり、それが続き、そして終わるといふ流れから、「開始」「継続」「終結」を設定できるという点では、概ね共通した見解が存在する。本研究では、松岡(他)(2009)およびTamaoka, et al. (2005)の研究成果を踏まえ、「開始」「継続」「終結」に、「状態」を加え、韓国語に対する分析を行う。

### 3. 高頻度の2字漢字語2,000語の抽出

本研究では、調査の基本データを得るべく、『21世紀世宗計画<sup>4</sup>』コーパスに対する語彙抽出を行った。『21世紀世宗計画』コーパスは、現代韓国語の話し言葉と書き言葉を収録した均衡コーパスである。語彙抽出は、次の手順で行った。(1)『21世紀世宗計画』コーパス全体を形態素解析して、(2)正規表現を使い、漢字2つで構成される漢字語(表記はハングル)を収集した。そして、(3)使用頻度が高い順に並び替え、上位2,000語を選んだ。以下では、この2,000語を「基本資料語」と呼ぶ。

「基本資料語」に対して次の処理を行った。韓国の国語辞典である『標準国語大辞典』を参考にして、漢字に変換した。なお、同音異義語であることが確認された場合には、コーパスに掲載された実際の用例150文を目視で確認し、文脈に合う漢字表記に変換した。

さらに、抽出した2,000語の妥当性を検討すべく、コーパス全体の語彙量を調査した。まず『21世紀世宗計画』コーパスの延べ頻度は26,329,643語であり、本研究の「基本資料語」の延べ頻度は1,481,916語である。その全体に占める割合は、5.63% (1,481,916/26,329,643)である。また、コーパス全体の名詞の延べ頻度は5,137,556語であり、この「基本資料語」の名詞全体に占める割合は28.84% (1,481,916/5,137,556)である。以上の結果から、本研究の「基本資料語」は、コーパス全体の語彙量から見ても、動詞化あるいは形容詞化を検討するためには十分なデータ量であると言える。

### 4. 2字漢字語の語彙性アスペクトの判定基準

本節では、「基本資料語」の語彙性アスペクトを判定する基準について述べる。「開始」「継続」「終結」「状態」の順に述べる。

#### 4.1. 「開始」のアスペクトの判断基準

松岡(2005)は、「～以内、～まで、～で」を使用して「開始」のアスペクト

<sup>4</sup> [http://www.sejong.or.kr] 韓国政府の文化観光部が、国立国語院および関連学会と共に、国語情報化発展計画の一環として、1998年から推進してきた国語情報化事業である。

を確認するテストを行った。例えば、「歯の治療後1時間以内の食事」であれば、“歯を治療した後、1時間以内に食べ終わるような食事”という意味より、“歯を治療した後、1時間以内に食べ始める食事”と解釈される。このことから、「以内」と共起する漢字語動詞は、開始の動詞性アスペクトを持つと述べている。松岡(2005)の基準は韓国語においても同様に適用できる。論証として、韓国語でも始めを表す表現があり、それが漢字語に後続することで開始を表すことができる。例えば、「식사시작(食事始作)/sigsasijag/」のように、韓国語の漢字語に「-시작(始作)/sijag/」(日本語の「始め」に相当する)を後続して使える場合は、「開始」のアスペクトを持っていると判断した。また、「瞬間的な意味」を持つ語や状態性を表す語は「開始」を持たないと判断した。

#### 4.2. 「継続」のアスペクト判断基準

松岡(2004)は、例えば「新しいビルが建設中だ」という文を作り、「-中」の接尾辞が共起する場合は「継続」のアスペクトを表すと判断している。韓国語でも、日本語と同様に「-중(中)/-jung/」という接尾辞があり、「건설중(建設中)/geonseoljung/」は「建設」という出来事が継続していることを示す。同様の事例として、「식사중(食事中)/sigsajung/」は、「食事」が継続していることを表す。以上の言語事実から「-중(中)/-jung/」との共起可能性に基づいて「継続」アスペクトかどうかを判断した。

#### 4.3. 「終結」のアスペクト判断基準

Iida(1987)およびShibatani & Kageyama(1988)は、出来事を表す名詞の中に、「一直後」などの時/アスペクトを表す接辞あるいは名詞がつけられるものがあると指摘しているが、松岡(2004)では、このアスペクト接尾辞の共起可能性に基づいて名詞の終結性を判断し、「終結」のアスペクトの有無を検討している。韓国語においても、日本語と同様に「식사후(食事後)/sigsahu/」のように「-후(後)/-hu/」を付けて瞬間および完了の意味を表すことができる。以上の言語事実から、「-후(後)/-hu/」との共起可能性に基づいて「終結」のアスペクトかどうかを判断した。

#### 4.4. 「状態」のアスペクト判断基準

松岡(2004)は、ある状態の持続時間を限定するということは、状態性を帯びた名詞であっても、時間的変化の可能性があると述べている。そして、名詞に「一時的な」という持続時間を限定する修飾句を加えた時に、その名詞が状態の一時的持続を表すかどうかで、「状態」のアスペクトの有無を判断できると述べている。韓国語でも、「일시적인 대립(一時的な対立)/ilsijeog-in daelib/」のような例があり、「일시적인(一時的な)/ilsijeog-in/」による修飾可能性を確認することで、「状態」のアスペクトかどうかを判断できる。さらに、韓国語では、「영구적인(永久的な)/yeonggujeog-in/」と「순간적인(瞬間的な)

/sunganjeog-in/」のような時間的な属性を持った修飾句を加えることもできるので、これらも「状態」のアスペクトの判定指標として加えた。

## 5. データ入力と信頼性

### 5.1. *-hada* 付加の辞書記述と韓国語母語話者の判定の信頼性

「基本資料語」に、*-hada* が付加されて動詞または形容詞になるかどうかを、次の2つの方法で検討した。(1)『標準国語大辞典』で、漢字語が見出し語として掲載されている場合には、まず名詞についての説明がくる。続いて、動詞または形容詞としての説明が続く。したがって、辞書の記載を見れば、*-hada* 付加の可否が容易に判定できる。また、『標準国語大辞典』では、品詞が見出し語ごとに示されているので、判断はさらに簡単である。この基準で、*-hada* が付加されると記載されていれば1を、記載が無い場合は0を入力した。(2) 日常的な使用を確認するために、著者を含む2名の韓国語母語話者が主観判定を行った。「基本資料語」について日常的に *-hada* を付加して使う場合には1を、使わない場合には0を入力した。2名の判定者の一致度を検討したが、一致しなかったのは、2,000語の内78語(3.90%)のみであった。そして、判定者間信頼性を検討するため、Krippendorffの  $\alpha$  ( $\alpha$ ) の信頼性指標 (Hayes & Krippendorff, 2007) を使用し、一致度を調べた。その結果、単純一致率で、動詞の場合は0.984、形容詞の場合は0.937であった。Krippendorffの  $\alpha$  の信頼度係数も0.924 (10,000回のブートストラッピングの結果、95%信頼区間は  $\alpha=0.908$  から0.941) と極めて高かった。

次に、主観判定の結果と辞書の記述の一致度を調べた。同じく Krippendorff の  $\alpha$  ( $\alpha$ ) の信頼性指標を使用し、計算した結果、辞書と第1判定者間の単純一致率は、動詞は0.938で、形容詞は0.839であった。この主観判定の結果と辞書間の Krippendorff の  $\alpha$  の信頼度係数は、0.884 (10,000回のブートストラッピングの結果、95%信頼区間は、 $\alpha=0.863$  から0.903) と高かった。また、辞書と第2判定者間の単純一致率で動詞は0.896で、形容詞は0.811であった。さらに辞書と第2判定者間の Krippendorff の  $\alpha$  の信頼度係数は、0.845 (10,000回のブートストラッピングの結果、95%信頼区間は  $\alpha=0.821$  から0.865) と高かった。したがって、*-hada* 付加の辞書を基準とした判定は、韓国語母語話者の内省と照合して信頼できると言えよう。

### 5.2. 使用頻度と語彙性アスペクトの判定と信頼性

「基本資料語」の頻度は、『21世紀世宗計画』コーパスで出現する延べ頻度をそのまま使用した。さらに、「基本資料語」を対象に「開始」「継続」「終結」「状態」の語彙性アスペクトを持つかどうかを、韓国語母語話者(第1判定者)が判定した。4節に示した判断基準に基づき、「基本資料語」について、「開始」「継続」「終結」「状態」のそれぞれのアスペクトを持っていれば1を、持っていない場合は0を入力した。以上の方法で判断した結果を、上位20位まで示したのが

表1である。これら 20 語をみると、動詞化されるのは「始作」「政治」「生活」「教育」「意味」の 5 語で、形容詞化されるのは「必要」の 1 語のみであった。「始作」「政治」「生活」「教育」は何らかのアスペクトを持っているのに対して、「意味」はいずれのアスペクトも持っていない。表1から、語彙性アスペクトを持っていない語でも、*-hada* が付加されて動詞化あるいは形容詞化され得ることが分かる。

4 つのアスペクトの判断において、第 1 判定者の信頼性を検討するために、もう 1 人の韓国語母語話者を第 2 判定者として設定した。第 2 判定者には「基本資料語」からランダムに 200 語を選び、ブラインド方式で語彙性アスペクトの判定を依頼した。両判定者の単純一致率は 0.871 であった。さらに、Krippendorff の alpha の信頼性指標は、0.830 (10,000 回のブートストラップの結果、95% 信頼区間は、 $\alpha=0.782$  から 0.876) とかなり高かった。したがって、本研究の第 1 判定者による「基本資料語」の語彙性アスペクトの判定は信頼性が高いと言えよう。

表1. 「基本資料語」の上位20語の2字漢字語

#	韓国語	ローマ字表記	漢字表記	使用頻度	<i>hada</i> 付加		開始 (inchoative)	継続 (durative)	終結 (telic)	状態 (stative)
					形容詞	動詞				
1	문제	munje	問題	12877	0	0	0	0	0	0
2	사회	sahoe	社會	11763	0	0	0	0	0	0
3	사실	sasil	事實	9017	0	0	0	0	0	0
4	시작	sijig	始作	7898	0	1	0	0	1	0
5	인간	ingan	人間	7536	0	0	0	0	0	0
6	정도	jeongdo	程度	7131	0	0	0	0	0	0
7	필요	pil-yo	必要	7108	1	0	0	0	0	1
8	정부	jeongbu	政府	6671	0	0	0	0	0	0
9	문화	munhwa	文化	6006	0	0	0	0	0	0
10	정치	jeongchi	政治	5901	0	1	1	1	1	0
11	의미	uimi	意味	5733	0	1	0	0	0	0
12	시간	sigan	時間	5727	0	0	0	0	0	0
13	여자	yeoja	女子	5559	0	0	0	0	0	0
14	국민	gugmin	國民	5507	0	0	0	0	0	0
15	경제	gyeongje	經濟	5492	0	0	0	0	0	0
16	대학	daehag	大學	5484	0	0	0	0	0	0
17	세계	segye	世界	5336	0	0	0	0	0	0
18	학교	haggyo	學校	5258	0	0	0	0	0	0
19	생활	saenghwal	生活	5183	0	1	1	1	1	0
20	교육	gyo-yug	教育	5095	0	1	1	1	1	0

## 6. 分析および分析結果と考察

### 6.1. 二項ロジスティック回帰分析

二項ロジスティック回帰分析は、2 値しかとりえない質的なデータを予測する手法である。ある事象の発生率を  $p$  とすると、オッズは  $p/(1-p)$  であり、その対数をとった  $\text{Log}(p/(1-p))$  が対数オッズと呼ばれる。ロジスティック回帰モデルは、 $\text{Log } p/(1-p) = B_0 + B_1X_1 + B_2X_2 + \dots + B_pX_p$  の式で表される。これは、あ

る事象がおこる発生率である目的変数を、 $p$  個の説明変数で予測する式である。Wald 統計量で各目的変数の貢献度を示す。本研究の場合は、2 字漢字語に *-hada* が付加される場合を 1、されない場合を 0 と記録して、この手法を適用した。具体的には、「基本資料語」について、「開始」「継続」「終結」「状態」の 4 つの語彙性アスペクト ( $X_1$ - $X_4$ ) と使用頻度 ( $X_5$ ) の 5 つを説明変数とした。そして、「基本資料語」に *-hada* が付加されて動詞化または形容詞化されるかどうかを予測した。ただし、本研究では、*-hada* 付加の有無は、辞書によって決められているので、むしろ 4 つの語彙性アスペクトと使用頻度の 5 つの変数の予測力を、Wald 統計量を参照して検討している。

## 6.2. *-hada* 付加による動詞の予測

「基本資料語」の 2,000 語のうち、動詞として *-hada* が付加されるのは 843 語 (42.2%) であった。表 2 は、動詞としての *-hada* が付加される 843 語について、4 つのアスペクトの重なり頻度を示した。アスペクトごとの漢字語の数は、「開始」が 320 語、「継続」が 520 語、「終結」が 687 語、「状態」が 180 語である。また、「単独での語数」の欄で、各アスペクトだけしか持たない漢字語の数を示した。例えば、「開始」のアスペクトを持つ漢字語は 320 語で、このうち 287 語は「継続」、311 語は「終結」、22 語は「状態」のアスペクトも同時に持っている。「開始」のアスペクトだけしか持たない語は 0 であった。つまり、単独で「開始」のアスペクトだけを持つ漢字語は存在せず、すべてが「継続」か「終結」と重複することがわかる。

表 2. *-hada* が付加され動詞になる漢字語についての 4 種類のアスペクトの頻度

変数名	アスペクト	合計	重複する漢字語数				単独での語数
			開始	継続	終結	状態	
$X_1$	開始	320	/	287	311	22	0
$X_2$	継続	520	287	/	493	44	11
$X_3$	終結	687	311	493	/	83	120
$X_4$	状態	180	22	44	83	/	89

注 1：高頻度の 2 字漢字語の合計は、2,000 語である。

注 2：2,000 語のうち動詞になるのは 843 語であり、表 2 はその集計である。

注 3：灰色のセルは、その特定のアスペクトだけの頻度を示す。

注 4：「単独での語数」とは、特定のアスペクト 1 つしかない漢字語の数である。

「終結」のアスペクトを持つ漢字語は、特に「開始」と「継続」のアスペクトとの重複が多かった。具体的には、「終結」のアスペクトを持つ漢字語 687 語の

内、311 語が「開始」、493 語が「継続」のアスペクトを持っていた。これは、動作や行動の「開始」「継続」「終結」という一連の時間性が内在しているためである。また、「終結」だけのアスペクトを持つ漢字語も 120 語あった。これには、「出発」「到着」などの瞬間動詞が多く含まれていた。瞬間動詞には、動作や行動の時間相が存在しない。そのため、「開始」や「継続」のアスペクトと重複することがないので、「終結」のアスペクトのみを持つ漢字語が多くなったのであろう。また、「状態」のアスペクトは、時間的な相を含まないため、それだけのアスペクトしか持たない漢字語も 89 語と多かった。

強制投入法による二項ロジスティック回帰分析の結果、 $\text{Log } p/(1-p) = 1.216X_1 + 1.755X_2 + 4.518X_3 + 2.027X_4 + 0.000X_5 - 2.857$  の回帰式が得られた。まず、モデル係数のオムニバス検定の結果、 $\chi^2(5)=1,693.649, p<0.001$  で有意であり、この回帰式は動詞としての *-hada* 付加を予測するのに役立っていることを示している。回帰式の当てはまりの良さを示す決定係数は、Nagelkerke が  $R^2=0.768$ 、Cox-Snell が  $R^2=0.571$  で高かった。表 3 に示したように、動詞としての *-hada* 付加は、4 つのアスペクトすべてで有意に予測された。そのなかで、説明変数の予測力の指標である Wald 統計量が最も大きかったのは「終結」の 380.562 である。実際、動詞としての *-hada* が付加される 843 語のうち 687 語を予測しており、これは 81.49% の予測率になる。

もちろん、「継続」のアスペクトでも 843 語の内 520 語を予測し、61.68% の高い予測率になる。しかし、「継続」と「終結」は、493 語も重複しているので、「継続」の Wald 統計量は 35.677 と小さくなり、予測力は「終結」に比べてかなり弱い。一方、「状態」のアスペクトについては、動詞としての *-hada* が付加される漢字語の 843 語の内 180 語を予測し、それは 21.35% (180/843) の予測率でしかない。一見すると、予測力が弱そうに見える。しかし、「状態」のアスペクトを持つ 180 語の内、「終結」と共有するのは、わずかに 83 語であり、46.11% (83/180) の共有率であった。そのため、「終結」のアスペクトとの重複が少ないので、表 3 の Wald 統計量を見ると、120.99 と高く、予測力は、「終結」に次いで強かった。使用頻度は、*-hada* 付加をまったく予測しなかった。日本語の場合 (松岡他, 2009; Tamaoka, et al., 2005) と同様に、韓国語の漢字語の 2,000 語においても、「終結」のアスペクトだけで動詞としての *-hada* 付加が 843 語のうち 687 語、つまり 81.49% を予測できるという結果であった。

表3. *-hada*付加(Y)による動詞の4種類のアスペクト(Xn)で予測する二項ロジスティック回帰分析の結果

変数名	アスペクト	アスペクト 含有数	アスペクト 動詞としての <i>-hada</i> 付加・非付加語数	アスペクト内 の予測率	全 <i>-hada</i> 付加語 内の予測率	二項ロジスティック回帰分析の結果			
						B	Wald統計量	有意確率	オッズ比
X <sub>1</sub>	開始 (inchoative)	330	付加 320	96.970%	37.960%	1.197	7.393	0.007	3.311
			非付加 10						
X <sub>2</sub>	継続 (durative)	554	付加 520	93.863%	61.684%	1.746	35.677	0.000	5.734
			非付加 34						
X <sub>3</sub>	終結 (telic)	720	付加 687	95.417%	81.495%	4.525	380.562	0.000	92.250
			非付加 33						
X <sub>4</sub>	状態 (stative)	386	付加 180	46.632%	21.352%	2.030	120.992	0.000	7.612
			非付加 206						
X <sub>5</sub>	使用頻度 (frequency)	—	付加 —	—	—	0.000	0.032	0.859	1.000
			非付加 —						

仮に、「終結」の  $X_3$  と「状態」の  $X_4$  のアスペクトで動詞化が予測できるのであれば、この2つだけを予測変数とした分析で、予測力を比較することができる。強制投入法による二項ロジスティック回帰分析の結果、 $\text{Log } p/(1-p) = 5.614X_3 + 1.906X_4 + 0.000X_5 - 2.691$  の回帰式が得られた。モデル係数のオムニバス検定の結果、 $\chi^2(2) = 1,623.953, p < 0.001$  で有意であり、回帰式が予測するのに役立っていた。回帰式の決定係数は、Nagelkerke が  $R^2 = 0.748$ 、Cox-Snell が  $R^2 = 0.556$  で、動詞化の予測への貢献度が高かった。また、2つのアスペクトの個々の Wald 統計量は、「終結」が 654.77 で、「状態」が 113.627 となり、両者とも非常に高かった。このように、4種類の語彙性アスペクトすべてで予測した場合と、2種類のアスペクトで予測した場合の違いは、わずかであり、「終結」と「状態」で動詞化がほとんど予測できることが分かる。

### 6.3. *-hada* 付加による形容詞の予測

漢字語 2,000 語について、形容詞としての *-hada* が付加されるかどうかと 4つのアスペクトの重なりを示したのが表4である。2,000 語のうち、形容詞としての *-hada* が付加されるのは 86 語で、わずかに全体の 4.30% に過ぎない。表4は、形容詞の 86 語が持っている4つのアスペクトの重複語数を示している。形容詞としての *-hada* が付加されるのは、「状態」のアスペクトを持つ漢字語が 78 語と非常に多く、重複は「終結」のアスペクトで 2 語しかなかった。「状態」のアスペクトが形容詞としての *-hada* の付加を強く予測すると想定される。

動詞の場合と同様に、強制投入法による二項ロジスティック回帰分析を行った。その結果、 $\text{Log } p/(1-p) = 0.552X_1 - 2.037X_2 - 1.651X_3 + 3.777X_4 + 0.000X_5 - 4.772$  の回帰式が得られた。モデル係数のオムニバス検定の結果、 $\chi^2(5) = 253.603, p < 0.001$  で有意であり、この回帰式は形容詞としての *-hada* 付加を予測するのに役立っている。回帰式の当てはまりの良さを示す決定係数は、Nagelkerke が  $R^2 = 0.399$ 、Cox-Snell が  $R^2 = 0.119$  であまり高くなかった。説明力が低くなったのは、2,000 語に占める形容詞の語数が少なかったことによる。表5に示したように、形容詞としての *-hada* の付加は、「状態」と「終結」のアスペクトによって有意に予測できた。「状態」は、Wald 統計量が 97.043 ( $p < 0.001$ ) で、形容詞としての *-hada* が付加される 86 語のうち 78 語が「状態」のアスペクトで予測し、予測率は 90.70% になる。

表4. *-hada*が付加され形容詞になる漢字語についての4種類のアスペクトの頻度

変数名	アスペクト	合計	重複する漢字語数				単独での語数
			開始	継続	終結	状態	
X <sub>1</sub>	開始	1		1	1	0	0
X <sub>2</sub>	継続	1	1		1	0	0
X <sub>3</sub>	終結	4	1	1		2	1
X <sub>4</sub>	状態	78	0	0	2		76

注1：高頻度の2字漢字語の合計は、2,000語である。

注2：2,000語のうち形容詞になるのは86語であり、表4はその集計である。

注3：灰色のセルは、その特定のアスペクトだけの頻度を示す。

注4：「単独での語数」とは、特定のアスペクト1つしかない漢字語の数である。

「終結」のアスペクトの、Wald 統計量が8.014 ( $p < 0.05$ )で、有意に形容詞としての *-hada* の付加を予測する。しかし、表4から分かるように、これに該当するのはわずかに4語である。さらに、そのうちの2語は「状態」のアスペクトを共有している。そのため、Wald 統計量の値も小さく、予測力は極めて弱いと言えよう。実際、386語が「状態」のアスペクトを持つものの、この内の *-hada* が付加されて形容詞になるのはわずかに86語であり、その内の78語を予測したに過ぎない。つまり、「状態」のアスペクトを持つからといって、すべてが形容詞化されるとはかぎらない。

#### 6.4. 動詞化および形容詞化と「状態」のアスペクトの関係

2,000語のうち「状態」アスペクトを持つ漢字語は、386語である。これらの内、*-hada* を付加されることで動詞になるのは、180語である。一方、形容詞になるのは78語である。「状態」のアスペクトを持つ漢字語で、動詞化される語数は、形容詞化される語数の2倍以上になる。しかし、「状態」のアスペクトを持つ漢字語の386語でみると、「状態」のアスペクトを持つ漢字語で動詞化されるのは46.63%、形容詞化されるのは20.21%になる。さらに、「状態」のアスペクトのみで動詞化するのは89語(23.06%)であり、形容詞化するのは76語(19.69%)である。つまり、「状態」のアスペクトを持っている漢字語であっても、動詞化あるいは形容詞化するわけではないことが分かる。

表5. *-hada* 付加(Y)による形容詞の4種類のアスペクト(Xn)で予測する二項ロジスティック回帰分析の結果

変数名	アスペクト	アスペクト 含有数	アスペクト としての <i>-hada</i> 付加・非付加語数	アスペクト 内の予測率	全 <i>-hada</i> 付加語 内の予測率	二項ロジスティック回帰分析の結果				
						B	Wald統計量	有意確率	オッズ比	
X <sub>1</sub>	開始 (inchoative)	330	付加	1	0.303%	1.163%	0.423	0.121	0.728	1.527
			非付加	329						
X <sub>2</sub>	継続 (durative)	554	付加	1	0.181%	1.163%	-1.980	3.179	0.075	0.138
			非付加	553						
X <sub>3</sub>	終結 (telic)	720	付加	4	0.556%	4.651%	-1.677	8.014	0.005	0.187
			非付加	716						
X <sub>4</sub>	状態 (stative)	386	付加	78	20.207%	90.698%	3.739	97.043	0.000	42.048
			非付加	308						
X <sub>5</sub>	使用頻度 (frequency)	—	付加	—	—	—	0.000	0.132	0.717	1.000
			非付加	—						

## 7. 総合考察

本研究では、韓国語の *-hada* 付加による動詞化と形容詞化について、2 字漢字語の語彙性アスペクトの予測力を検討した。本研究では、Tamaoka et al. (2005) および松岡他 (2009) の二項ロジスティック回帰の手法による日本語のサ変動詞 *-suru* 付加の研究を適用した。そして、韓国語の 2 字漢字語のうち使用頻度が高い 2,000 語について、「開始」「継続」「終結」「状態」の 4 つのアスペクトと使用頻度から *-hada* 付加を動詞化と形容詞化を別々に集計して予測した。その結果、2 字漢字語の使用頻度は、*-hada* 付加による動詞化または形容詞化をまったく予測しなかった。つまり、漢字語がどのくらい頻繁に使われるかは無関係であることが分かった。

2,000 語のうち、動詞としての *-hada* が付加されるのは 843 語 (42.15%) であった。4 つのアスペクトの内、*-hada* 付加による動詞化を最もよく予測したのは「終結」のアスペクトであり、Wald 統計量が 380.56 となり、予測力が非常に強かった。具体的には、843 語のうち 687 語を予測し、予測率は単独で 81.49% になった。また、「継続」のアスペクトを持つ漢字語は 554 語で、843 語の内 520 語を予測し、その予測率は 61.68% であった。しかし、「終結」との重複が 493 語もあるため、4 つのアスペクトを合わせた場合の「継続」の予測力はやや低かった (Wald 統計量=35.68)。さらに、「開始」のアスペクトの予測率は 37.96% (320/843) であるが、他のアスペクトと重複しているため、Wald 統計量は 7.39 で、予測力はさらに低かった。つまり、「開始」のアスペクトは無くても、「終結」または「継続」で代用できるということになる。一方、「状態」のアスペクトについては、*-hada* が付加されて動詞になる漢字語の 843 語の内 180 語を予測し、予測率は 21.35% (180/843) になった。予測率は低いので一見すると予測力も弱そうに思えるが、他のアスペクトとの重複が少ないため、Wald 統計量は 120.99 となり、予測力は「終結」に次いで強かった。

以上のように、*-hada* 付加による動詞化の予測は、まず「±終結」のアスペクトによる予測力が極めて強いと言えよう。表 2 からアスペクトの重複を見ると、「開始」のアスペクトを持つ 320 語の内、「終結」のアスペクトを共有するのは 311 語であり、97.19% (311/320) である。また、「継続」については、520 語の内 493 語を共有しており、94.81% (493/520) に当たる。一方、「状態」との共有は、180 語の内 83 語であり、46.11% (83/180) であった。そのため、大多数の「開始」と「継続」のアスペクトを持つ 2 字漢字語が「終結」のアスペクトを持っていることになり、「終結」だけでほとんどの 2 字漢字語の動詞化を予測できることになる。一方、「状態」については、「終結」との重複が半分以下であり、予測力はある程度強くなった。さらに、「終結」と「状態」のアスペクトだけで動詞化を予測しても、4 種類全部で予測した場合とあまり変わらなかった。そのため、韓国語の 2 字漢字語の動詞化は、大多数が「±終結」で、次いで「±状態」のアスペクトの特性によってほぼ決まると言えよう。

一方、2,000 語のうち *-hada* が付加されて形容詞になるのは 86 語で、わずか

に 4.30%に過ぎない。それを有意に予測したのは、「状態」と「終結」のアスペクトであった。「状態」は、全 86 語の内 78 語を予測し、90.70%の予測率と非常に高かった。「終結」は、わずか 4 語の漢字語にしかみられず、そのうち 2 語は「状態」のアスペクトと共有している。したがって、*-hada* の付加で形容詞になるのは、ほぼすべてにおいて「±状態」のアスペクトの特性を持っていると言えよう。なお、「終結」と「状態」のアスペクトが重複するのは、「만족 (満足) /manjog/」と「충족 (充足) /chungjog/」の 2 語のみである。これらの語は、動詞にも形容詞にもなりうる。

「状態」のアスペクトが形容詞化をほぼ予測するように思えるが、予測力指標である Wald 統計量は 97.04 で、動詞化の場合と比べてずっと低い。これは「状態」のアスペクトを持つ 2 字漢字語の 2,000 語の内、386 語が「状態」のアスペクトを持つものの、その内、形容詞化するのはわずかに 78 語であることが原因である。つまり、「状態」のアスペクト全体で見ると、形容詞化の予測率は 20.21% (78/386) に過ぎないのである。確かに、「状態」のアスペクトは形容詞化を予測する必要条件ではある。しかし、むしろ形容詞化しない場合の方が多く、十分条件というには程遠い。つまり、「努力 (노력/nolyeog/)」「認識 (인식/insig/)」「強調 (강조/gangjo/)」「記憶 (기억/gi-eog/)」「独立 (독립/doglib/)」など大多数の 79.79% (100%-20.21%) の「状態」のアスペクトを持つ 2 字漢字語は形容詞化しないのである。

## 8. おわりに

本研究では、漢字 2 字で表記されうる韓国語の上位 2,000 の高頻度語を対象に、辞書で判定された動詞化または形容詞化を、「開始」「継続」「終結」「状態」の 4 つの語彙性アスペクトと使用頻度で予測した。動詞化を強く予測したのは、「終結」のアスペクトであり、それに次いで、「終結」との重複が比較的少ない「状態」のアスペクトであった。一方、形容詞化には、「状態」のアスペクトの予測率が極めて高かった。しかし、形容詞化された 2 字漢字語はほぼすべて「状態」のアスペクトを持つものの、「状態」のアスペクトを持つ 2 字漢字語は、ほとんど形容詞化されない。そのため、予測力を示す Wald 統計量は、「終結」の動詞化と比べて、「状態」の形容詞化の Wald 統計量はかなり小さかった。

## 引用文献

- 金田一春彦 (1976). 「国語動詞の一分類」『日本語動詞のアスペクト』東京：むぎ書房, pp. 5-26.
- 平嶋一美 (2003). 「フランス語動詞事象の意味分類に関する考察」『外国語教育研究』5: 55-67.
- 松岡知津子 (2004). 「複合動詞「～スル」を形成する漢語名詞について」『日本語教育』120: 13-22.
- 松岡知津子 (2005). 『機能動詞スルと名詞が形成する文の構造と意味—漢語名

- 詞との結びつきを中心に』広島大学大学院教育研究科博士論文。
- 松岡知津子・玉岡賀津雄・酒井弘 (2009). 「アスペクトによる漢字二字熟語のサ変複合動詞化に対する予測」由本陽子・岸本秀樹 (編) 『語彙の意味と文法』東京:くろしお出版, pp. 121-137.
- 油谷幸利 (1978). 「現代韓国語動詞分類」『朝鮮学報』 87: 1-35.
- 国立国語院 (1999). 『標準国語大辞典』 Doosan Dong-a, [http://www.korean.go.kr/09\\_new/index.jsp](http://www.korean.go.kr/09_new/index.jsp) (2010年3月10日に閲覧)
- 国立国語院(2007). 『21世紀世宗計画』韓国観光文化部, <http://www.sejong.or.kr>. (2010年3月10日に閲覧)
- 李鷹百 (1980). 「국어辞典 語彙의 類別構成비교 漢字語의 重要度와 教育問題」 『語文研究』 8: 136-137.
- Hayes, A. F. & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1), 77-89.
- Iida, M. (1987). Case-assignment by nominals in Japanese. In M. Iida, S. Wechsler & D. Zec (Eds.), *Working papers in grammatical theory and discourse structure: Interactions of morphology, syntax, and discourse* (pp. 93-138). Stanford, CA: CSLI.
- Shibatani, M. & Kageyama, T. (1988). Word formation in a modular theory of grammar: Postsyntactic compounds in Japanese. *Language*, 64(3), 451-484.
- Tamaoka, K., Matsuoka, C., Sakai, H., & Makioka, S. (2005). Predicting attachment of the light verb-sure to Japanese two-kanji compound words using four aspects. *Glottometrics*, 10, 73-81.
- Yokosawa, K. & Umeda, M. (1988). Processes in human Kanji-word recognition. *Proceedings of the 1988 IEEE international conference on systems, man, and cybernetics*, 377-380. August 8-12, Beijing and Shenyang, China.
- Vendler, Z. (1967). *Linguistics in philosophy*. Ithaca, NY: Cornell University Press.

## **Predicting verbalization and adjectivization of Sino-Korean two-character compound words by four lexical aspects**

Seon-Ju PARK, Nagoya University

Katsuo TAMAOKA, Nagoya University

Jae-Ho LEE, University of Tsukuba

### **Abstract**

The top 2,000 high-frequency Sino-Korean words, composed of two Chinese characters, were taken from the Balanced Corpus of Korean. A logistic regression analysis was conducted to predict verbalization and adjectivization of these words by the four lexical aspects of *stative*, *telic*, *durative* and *inchoative*. The results showed significant predictions of *telic* and, to a lesser degree of *stative* for verbalization, and *stative* for adjectivization.