

形態的情報を使った日本語辞書情報の整備

6F-4

荻野紫穂

日本アイ・ビー・エム株式会社 東京基礎研究所

1. はじめに

大量のエントリに対して多数の情報が付加されている解析用辞書などにおいては、その作成や変更の際に、対象としている語の形態的・意味的派生関係をたどれると、情報の付加・統一の面で役立つことが多い。動詞に関しては、いわゆる自動詞・他動詞の派生関係を見ることで、態の変換による格交替などの情報が得られる。しかし、自動詞・他動詞の派生対応は、一律に揃った形をしている訳ではなく、規則化もしにくいので、大量のエントリに派生関係を付けるのは、かなり困難な作業となる。

本論文では、動詞の自他派生のモデルをもとに、辞書に登録されている動詞の自他派生関係を推測し情報を抽出する方法を述べる。

2. 動詞の自他派生モデル

2.1. 使用自他派生モデル

IPAL[4]には、自動詞・他動詞の派生語情報が登録されている。但し、登録語数が限られていることもあり、派生関係が挙げられていないものもしばしば見られる。動詞の自他の派生関係については、水谷[1]、村木[7]などに詳しい対応表が挙げられているが、本論文では、それらを単純・形式化した星野[5]の派生モデルを使用する。

このモデルに、派生の関係付けと情報抽出の観点から多少変更を加える。星野のモデルでは、-iru、-eruは、自動詞・他動詞双方の-uから派生することになっている。細かく見ると、-iru、-eru型の動詞は、大抵が、いわゆる古語の上二段動詞・下二段活用動詞が口語化したものである。二段活用の終止形は-uなので、-iru、-eruは、自分の属する自/他動詞側の-uが変形したものと大雑把に推測することができる。但し、二段活用動詞から変化した一段活用動詞-iru、-eruに対して、-u型の動詞が残っていないということは、原則的には保証はされない。

-iru、-eru以外の自動詞は、違う意味で使える表現を増やすという意味で、派生する場合には、自分とは違う側の-uから派生した、と考えるのがもっともらしい。他動詞方向の派生の場合は、特に使役的派生は、自動詞・他動詞の区別に関わらず起こり得るので、自動詞・他動詞双方の-uから派生した可能性がある。これらの点を考慮に入れたものが、図1の改訂モデルである。

2.2. 派生モデルの例外

水谷[1]や村木[7]には、星野のモデルで扱われる以外の自他の対立、例えば、-oreru (ex. 埋もれる)の自動詞化や-ku (ex. 破く)などの他動詞化が挙げられている。派生の関係をできるだけ表層の情報で判断できるようにするため、本

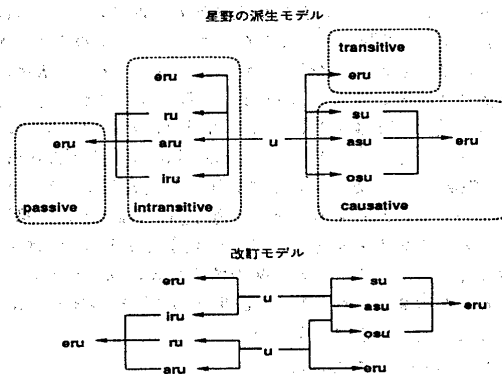


図 1: 星野の派生モデルと改訂モデル

論文では、情報をできるだけ表層の部分から得ようとするため、モデル化が明確になっていないこれらの型はデータの範囲から除いた。また、星野のモデルでは、-kasu (ex. 脅かす)は-asuの変形としてデータの範囲内としているが、同様の理由で、データから除いた。派生変形を情報源とするため、自他同形のものしかない派生型も、範囲外とする。

3. 情報抽出への適用

3.1. データの整理

前章で述べたモデルに実際の動詞を当てはめた結果を、表1に示す。初期データには、村木[7]の自他対応の型で示された動詞のうち、派生モデルに該当する動詞140を使用した。表中、「*」で埋められている欄が、その派生型の存在を示す。「+」はヲを格フレームに持つ型、つまりいわゆる他動詞と呼ばれるものであるにも関わらず、自動詞側に分類されたものを示す。

-ru、-suは、(その前に母音を指定していないので)-uとも考えられる場合もあるが、-ru、-suを取り去った部分に、他の派生型を直接接続したものが存在すれば、それを-ru、-su

	reru	iru	aru	ru	eru	iu	u	eru	su	asu	osu	seru
あく						*		*				
裂く					*		*					
減ぶ		*				*						*
帰す				*					*			
繋ぐ			*				*	*				
預る			+					*				
寝る				*								*
着る				+								*
煮る				+	*							

表 1: 派生モデル適用例

として分類する。

3.2. 可能の派生と使役・受身の派生

実在する辞書、特にユーザ辞書などの場合は、よく使う語を頻繁に登録している可能性があるため、このモデルを適用する場合に、多少の混乱を招く場合があるエントリがある。

一つは可能の派生である。五段活用動詞には、意味的制約がない場合には、-u、-eru という可能形の派生がほぼ全てにあり得る。これらが辞書に登録されている場合には、-eru 型の自動詞・他動詞と同形になりやすく、-u 型が現代語にある場合には、可能か自／他動詞かを見分けることは、表層的には困難である。更に、-reru は、五段活用の受身ともなるが、-u が存在する場合に、これも見分けがつけにくい。

もう一つは使役派生である。これは、派生モデルでも -asu、-seru で扱っているが、どこまでを派生と見るかに加え、「溶かす」のように現存する -u 型（「溶く」）とはほとんど格関係が変わらないものもあるので、注意を要する（「溶ける」の原型「溶く」と比べると、使役的格交替が見られる）。

これらを考慮に入れた上で、派生語の語尾を処理し、表 1 のような自他の関連付けを行なう。

3.3. -eru・自他の判断

-eru は自動詞・他動詞双方の可能性があるので、辞書のエントリにモデルを当てはめる場合、他の派生関係を利用して、ある程度は推測することができる。

同じ -u について、二種の自動詞の派生語があるよりは、自動詞・他動詞の派生があると考える方が、語の使い分けという意味でもっともらしい。従って、同じ -u に対して -eru と他の自動詞派生、つまり -aru、-iru が存在する場合には、-eru は自動詞でなく他動詞である可能性が高い。但し、全てが必ず他動詞という訳ではなく、初期データ中 -eru、-aru または -iru を持つもの 25 について、<抜く、抜かる、抜ける、抜かす><転がる、転げる、転がす> はこれに当てはまらない。

更に、初期データで見ると、-asu、-eru の双方がある場合には、-eru は自動詞であることが多い。ただし、-asu は原則的に自動詞・他動詞の区別なく派生するので、-aru、-iru が存在する場合には、そちらの条件が優先される。この条件には、該当する 21 語中、<抜く、抜かる、抜ける、抜かす><転がる、転げる、転がす> の 2 語がやはり当てはまらない。（水谷 [1] にもある通り、「抜く」には二つの派生の枝があるからと思われる。）

その他のものは自動的に自他の区別をするのは難しい。しかし、-u と -eru だけが存在する場合、古語の二段活用から派生した -eru のもととなる -u は、現代語では二段活用が一段活用に变化していることを考えると、原則として派生元ではなく、逆の側に属するものである可能性が大きい。従って、その -u に自他の情報があれば、-eru はその逆側と推測することができる。ただし、数は少ないが、<含む、含める> のように両方他動詞、という場合もあり得る。

また、前章で述べた通り、可能の -eru が存在する場合には、表層情報から可能・自／他の判断をすることは困難である。

3.4. -u・自他の判断

-u も -eru と同様、自他双方の可能性があるので、-u は -eru に比べ判断材料は多くない。ただ、-u の判断に関して、-iru と -eru との間に対照的な現象が見られる。

-u、-iru の双方が存在する場合、-iru は自動詞であるが、-u

は -iru と同じ自動詞となる。初期データ以外の動詞でも調べた結果、-u、-iru が双方存在するものが 7 語あったが、その全てが双方自動詞であった。-eru も -iru も二段活用から変化した一段活用動詞だが、-iru が存在する場合には -u は -iru と同じ自動詞であり、-eru が存在する場合には、-u は -eru と逆の側に属していることが多い。前節でも述べた通り、-u、-eru だけが存在するものは、それだけでは自他の判断が付けられないが、何らかの形で、どちらかに自他の区別がつけられれば、もう一方はその逆の性質を持っていると推測できる。

4. 問題点～格の交替と自他の対応

特に -ru と -su / -seru の対応関係で考えると、「見る」のようないわゆる他動詞が -ru の自動詞側に入ることがしばしばある。これは、-ru、-su / -seru 間で特に顕著だが、<預る、預ける><授かる、授ける> のように、-aru、-eru 間でも見つかる。どちらも、

- a 私が見る。
- a' 彼が私に見せる。
- b 私が子どもを預る。
- b' 彼が私に子どもを預ける

のように、主格、あるいは動作に対する観点の変化は見られるが、早津 [6] などにあるように、「ヲ格をとるものが他動詞」という観点では、どれも他動詞になる。従って、一概に「-aru だから自動詞＝ヲ格をとらない」といった判断はできない（但し、「預る」「授かる」は数少ない例であり、実際の作業で「ヲ格をとるか」とらないかを根拠にしたい場合は、例外扱いをしてもあまり影響はない）。こういった派生関係を調べる場合には、自動詞・他動詞という大まかな区別に加えて、格の変換などの観点も必要となろう。

それに関して、-ru、-eru という対応があるものについては、<見る、見える><煮る、煮える> という <他動詞、自動詞> のペアがある。星野 [5] もこれを特殊型としている。これらは今回はモデルに当てはめきれなかった例であり、実際の処理では注意を要する。

5. まとめ

日本語動詞の自他派生対応モデルを使って、派生関係の情報を抽出する方法を述べた。派生関係は、大まかに自動詞・他動詞という区別を考えるだけでなく、格の変換などの観点も踏まえて整理する必要がある。

文献

- [1] 水谷静夫 (1982) 現代語動詞の所謂自他の派生対立『計量国語学』vol. 13, no. 5, pp. 212-222.
- [2] 寺村秀夫 (1982) 動詞の自他一語彙的態の類型『日本語のシンタクスと意味』2, くろしお出版, vol. 1, pp. 303-317.
- [3] 外池滋生 (1985) 態を軸とした動詞分類試案『ソフトウェア文書のための日本語処理の研究』5, 情報処理振興事業協会, pp. 117-132.
- [4] 情報処理振興事業協会 (1987) 『計算機用日本語基本動詞辞書 IPAL(Basic Verbs) 解説篇・辞書篇』.
- [5] 星野和子 (1988) いわゆる自他派生型に基づく動詞分類『東京女子大学日本文学』72, pp. 134-144.
- [6] 早津恵美子 (1989) 有対他動詞と無対他動詞の意味上の分布『計量国語学』vol. 16, no. 8, pp. 353-364.
- [7] 村木新次郎 (1991) 『日本語動詞の諸相』ひつじ書房.