

英語表現を含む日本語表現に関する研究

村田 真樹^{*1} 金丸 敏幸^{*2} 仲本 康一郎^{*1} 小谷 克則^{*1} 井佐原 均^{*1}

通信総合研究所^{*1}/ 京都大学^{*2} (murata@crl.go.jp)

1 はじめに

最近、「ワールドカップ in 埼玉」のような日本語中に英語表現が挿入された表現をよく見かける。そこで、そういう表現の抽出を行なった。抽出は字種を利用し、日本語文字と英語文字の双方が含まれる表現を計算機で自動抽出する。この実験を毎日新聞と web データで行なった。また、「博覧会イン埼玉」のように、英語表現を片仮名で表記している場合の表現も抽出した。この種の表現は最近急に見かけるようになった表現で、このような表現を研究することは言語の発展過程を調べる言語学に資するものである [1, 2, 3]。

2 英語・片仮名表現の自動抽出

日本語中に英語表現が挿入された表現の抽出を毎日新聞 (1991-2002 年) と Web データ (約 2GB) で行なった。抽出の手順は、まず計算機で粗く抽出しその結果を手で整理することによって行なった。

まず、毎日新聞で実験を行なった。毎日新聞の実験では抽出は鍵括弧内の表現に限った。鍵括弧で囲まれる表現で、同じ表現は削り、その表現で、1) 日本語文字 (平仮名・片仮名・漢字) と英語文字を含むこと、2) 英語表現の前後は日本語文字か鍵括弧表現であること、3) 二文字目以降は小文字の英語表現であること、4) 二文字以上ある英語表現であることを満足する表現を自動抽出した。本研究では、抽出する英語表現は一単語に限った¹。これは複数の単語からなる英語表現は英語の句や文を引用しているにすぎず、日本語と英語の混ざった表現ではない場合が多いからである。この自動抽出した結果で英語表現が名詞、固有名詞、記号、「な」のつく形容詞、「する」がつく動詞、英語語順ではない動詞²を除いたものを年度ごとに頻度順に整理し、この結果を表 1 に示している。本稿で示す頻度はそういう表現を含む日本語表現の異なり数である。表の「前」や「名」等は品詞を意味し「句」は句を意味している。この集計は自動抽出した結果をすべて人手で細かく調査して行なった³。

¹ 空白を挟まず Welcometo と一語で表記するものは抽出した。

² これらは従来の日本語でもそれほど珍しくないで除いた。

³ “de” は英語ではなくフランス語であるが、日本語文中に現れる

次に Web データを用いて実験を行なった。毎日新聞とは異なりすべてのテキストを対象とした。同じ表現の行は削除した。「5) 文頭では英語表現は大文字であること」を新たな条件として用いた以外は毎日新聞での実験と同じ条件で自動抽出した。自動抽出した結果を手で整理した⁴結果を表 3 に示している。(全角半角の区別をしている。) ここではデータは品詞ごとに整理し、英語表現とその頻度と実際にデータにあった表現例を並べて表記している。例では対象となった英語表現の部分は“*”で略記している。表ではなるべく頻度の高いものから順に表示している。毎日新聞での整理と同様に名詞等を除いている。この整理ではデータ量が多かったため、頻度は 10 以上あったもののみを示している。また、各英語表現の頻度は自動抽出した個数を表示して人手では頻度を数え直していない。

最後に片仮名表現を含む日本語表現の抽出の実験を行なった。これも毎日新聞と Web で行なった。抽出は英語表現の場合と同様で字種を利用し、日本語文字を平仮名漢字文字、英語文字を片仮名文字として全く同じ方法で行なった⁵。この実験ではデータ量が多くすべてを手で調査することが困難であるので、毎日新聞と Web データでの実験で抽出した英語表現を片仮名表記したリストを作成し、このリストにあった片仮名表現のみを手で調査した。結果は表 2 と表 4 に示している。Web データでは頻度 1 以上のすべての結果を示しているが英語表現の実験と同じく頻度は自動抽出した個数である。

3 考察

表 1 の毎日新聞の年度ごとの調査により前置詞 in は古くから使われていることがわかる。また 1994 年以降になると in の使用頻度が急に大きくなっていることがわかる。Let's も古くから使われている。at, de, of, with, 現象としては興味深いのでこれも表に載せている。本研究では前記の品詞で除いた結果では“de”以外には英語以外の外国語はなかった。

⁴ この整理は最初の 50 例までを見て行なった。

⁵ 片仮名表現を片仮名表現で挟む日本語表現は抽出されないことに注意してほしい。片仮名表現を片仮名表現で挟む日本語表現は、抽出が困難であることと、長い英語句を片仮名で表記しただけの表現である場合が多いことにより、本研究では扱わないことにした。

from,by,to等の前置詞も頻度は少ないが比較的古くから使われていることがわかる。deは1998年から頻度が大きくなっている。

表3のようにWebデータを利用することで様々な品詞の様々な英語表現を抽出できた。名詞などの対象としなかった表現以外では、間投詞と助動詞を除く英語のほとんどすべての品詞を抽出できた。前置詞が最も多くまた様々な前置詞を抽出できた。次にorなどの接続詞が多く、次に形容詞が多く出現している。以降は句、副詞、冠詞、所有代名詞、接尾辞、動詞、疑問代名詞が並ぶ。Webでは表記は全角と半角があるが全体的に半角の方がよく用いられている。

表3の英語の前置詞、句、動詞を使った表現は「キャンペーン in 渋谷」「What's けいば」「Welcome えひめ」のように英語の語順になっていて日本語と語順が異なり興味深い。コンパクトな表現で重要なことを先に持ってくるには英語表現が便利でそういう理由で生まれた表現なのかもしれない。

表3と表4を使って英語表現での頻度と片仮名表現での頻度を見比べると面白いことがわかる。既に日本語表現として定着していそうな、副詞「アップ」や形容詞「デジタル」は片仮名表現で多く、まだ日本語表現として定着していなさそうな前置詞表現は比較的片仮名表現で少ない。つまり、日本語表現として定着していそうな表現は片仮名表現が多く用いられ、日本語表現としてまだ定着していなさそうな表現は英語表現が多く用いられる傾向があることがわかる。抽出した英語表現では形容詞と副詞を除く前置詞、接続詞、句、冠詞、所有代名詞、接尾辞、動詞、疑問代名詞では英語表現の方が多く、これらは日本語表現としてまだ定着しきっていない比較的新しい表現形式であると思われる。また、形容詞と副詞においては、形容詞new,open、副詞all,notは英語表現の方が多くこれらは形容詞と副詞の中でも特異な表現であると思われる。ただしWebでの頻度は当該品詞や当該条件以外の頻度も含まれているもので詳細は今後調査・再検討する必要がある。

表2の片仮名表現の年度ごとの調査を見てみよう。形容詞、副詞は古くから頻度も多い。これは上述の形容詞、副詞が他のものよりも定着しているという議論と一致する。動詞「ウェルカム」、冠詞「ザ」、接続詞「アンド」は少数ではあるが古くから使われている。前置詞「イン」は1993年に1つ、1995年に2つ出現し以降毎年1～3個ぐらい出現している。この「イン」は英語表現の表1と見比べると面白く、1993年と1995年の一年前の1992年と1994年で4個と21

個と英語表現inが多く用いられ以降毎年二桁の大きい出現がある。日本語表現「イン」は英語表現inが使われた時から少し遅れた一年後に出現をしているようである。これは興味深い。

最後に抽出できた英語表現について片仮名表現があるかgoogleを使って人手で調べてみた。since,dear,max,not,It's,postedbyで見つからずvs,with,for,to,next,all(副詞)で少なそうな感触を得た。これら以外の表現はある程度以上見つかった。deについては「ネット・デ・ショップ」という形式で多く出現していた。片仮名表現の利用については片仮名で表現したときの見やすさも影響していそうである。

4 おわりに

毎日新聞12年分とWebデータ(約2GB)を使って英語表現を含む日本語表現の抽出を行なった。毎日新聞での年度ごとの実験結果から前置詞inは古くから使われていることがわかった。Webデータを使った実験により様々な品詞の英語表現を抽出できた。英語の前置詞、動詞句を使った表現については日本語と語順の変わるものであり興味深い。片仮名表現の抽出も行ない、片仮名表現と比較することでどういう表現が新しい表現形式であるかどうかの議論も行なうことができた。

今後の展望としては文体論と混成語との関わりを調べてみたいと思っている。英語表現を含む日本語表現は主にタイトル中に含まれる。タイトルでは簡潔に述べる必要があり、そのタイトルの持つ文体的特徴との関係を調べてみたい。また、英語表現を含む日本語表現は日本語と英語の混ざった表現であり、混成語と何らかの関係がありそうである。ただし、日本人同士でのコミュニケーションのために発展したものであり、一般の混成語とは異なるもので、研究としてはさらに興味深い側面があると思う。混成語との関わりという観点でも今後調べてみたい。

参考文献

- [1] 伊藤雅光, ホップス系流行歌の語彙調査における外来語と外国語の判定基準, 計量国語学, Vol. 23, No. 2, (2001), pp. 110-131.
- [2] 吉田桃子, 望月源, 日本語流行歌中の外来語に関する調査, 言語処理学会第9回年次大会, (2003).
- [3] 村田真樹, 仲本康一郎, 小谷克則, 井佐原均, 英語表現を含む日本語表現の抽出, 計量国語学第四十七回大会, (2003).

表 1: 毎日新聞での英語表現の集計

1991年	1992年	1993年	1994年	1995年	1996年
in 前 2 Let's 句 1 at 前 1 de 前 1 vs 前 1	in 前 4 The 冠 1 Handy 形 1 New 形 1 of 前 1	Let's 句 1 in 前 1 vs 前 1 with 前 1	in 前 12 the 冠 1 my 所 1 by 前 1 de 前 1 to 前 1	in 前 21 and 接 1 From 前 1 by 前 1 from 前 1 to 前 1 with 前 1	in 前 15 Dear 形 3 That's 句 2 New 形 2 Let's 句 1 de 前 1 to 前 1 vs 前 1 with 前 1 Be 動 1
1997年	1998年	1999年	2000年	2001年	2002年
in 前 28 With 前 2 Dear 形 1 Jr 形 1 for 前 1 vs 前 1 with 前 1 Forever 副 1	in 前 21 de 前 6 The 冠 3 Let's 句 1 ing 接 1 By 前 1 for 前 1 to 前 1 with 前 1	in 前 22 de 前 3 My 所 2 vs 前 2 It's 句 1 New 形 1 ing 接 1 or 接 1 from 前 1	in 前 19 de 前 9 at 前 2 from 前 2 with 前 2 and 接 1 by 前 1 to 前 1	in 前 32 de 前 6 New 形 2 from 前 2 to 前 2 the 冠 1 That's 句 1 What's 句 1 Digital 形 1 can 助 1 De 前 1 With 前 1 for 前 1 vs 前 1 again 副 1	in 前 13 de 前 7 New 形 2 and 接 1 from 前 1 with 前 1

表 2: 毎日新聞での片仮名表現の集計

1991年	1992年	1993年	1994年	1995年	1996年
アップ 副 11 ニュー 形 7 グローバル 形 5 スーパー 形 3 デジタル 形 3 ジュニア 形 2 パーソナル 形 2 ベスト 形 2 ザ 冠 1 ハウツー 句 1 オープン 形 1 オール 形 1 ノー 形 1 ウエルカム 動 1	デジタル 形 11 スーパー 形 7 ニュー 形 7 アップ 副 7 パーソナル 形 6 オール 形 3 ジュニア 形 3 ノー 形 3 ベスト 形 3 オープン 形 2 グローバル 形 2 スペシャル 形 2 ハイパー 形 2 ハウツー 句 1 アンド 接 1 ウエルカム 動 1	アップ 副 21 スーパー 形 8 デジタル 形 6 ジュニア 形 4 ニュー 形 4 ザ 冠 3 オール 形 3 オンライン 形 2 グローバル 形 2 ザッツ 句 1 オープン 形 1 ハンディー 形 1 パーチャル 形 1 フリー 形 1 ベスト 形 1 イン 前 1 オンリー 副 1	アップ 副 22 デジタル 形 14 スーパー 形 12 オール 形 4 オンライン 形 3 グローバル 形 3 ジュニア 形 3 オープン 形 1 オフライン 形 1 ニュー 形 1 ノー 形 1 パーチャル 形 1 パーソナル 形 1 ベスト 形 1 ウエルカム 動 1 エンジョイ 動 1 アゲイン 副 1	オール 形 20 デジタル 形 12 アップ 副 12 スーパー 形 9 グローバル 形 5 オンライン 形 4 オープン 形 3 イン 前 2 イツ 句 1 ハウツー 句 1 レッツ 句 1 キッズ 形 1 ジュニア 形 1 スペシャル 形 1 ビッグ 形 1 ベスト 形 1 イング 接 1 ウエルカム 動 1 サポート 動 1	アップ 副 20 デジタル 形 19 グローバル 形 10 スーパー 形 8 オール 形 6 パーチャル 形 5 ニュー 形 4 オープン 形 2 オンライン 形 2 ジュニア 形 2 ビッグ 形 2 ベスト 形 2 イン 前 2 ザ 冠 1 アクティブ 形 1 クローズ 形 1 スペシャル 形 1 ハイパー 形 1 マイ 所 1
1997年	1998年	1999年	2000年	2001年	2002年
デジタル 形 42 アップ 副 22 グローバル 形 18 スーパー 形 14 オール 形 11 パーチャル 形 8 ニュー 形 5 オープン 形 2 オンライン 形 2 ビッグ 形 2 スペシャル 形 1 ノー 形 1 メイン 形 1 モバイル 形 1 イン 前 1	デジタル 形 27 グローバル 形 15 アップ 副 10 スーパー 形 9 オール 形 8 ニュー 形 6 パーチャル 形 3 モバイル 形 3 オンライン 形 2 キッズ 形 2 ジュニア 形 2 ノー 形 2 アクティブ 形 1 デジタル 形 1 ナイス 形 1 ベスト 形 1 マンスリー 形 1 イン 前 1 フォーエヴァー 副 1	デジタル 形 39 グローバル 形 31 スーパー 形 10 アップ 副 10 オール 形 9 オンライン 形 8 ジュニア 形 6 パーチャル 形 3 キッズ 形 2 ネクスト 形 2 マンスリー 形 2 モバイル 形 2 ニュー 形 1 ノー 形 1 ハッピー 形 1 ベスト 形 1 エンジョイ 動 1	デジタル 形 46 グローバル 形 23 スーパー 形 17 オール 形 10 アップ 副 8 モバイル 形 7 オンライン 形 5 パーチャル 形 4 ジュニア 形 3 ニュー 形 3 ハイパー 形 3 キッズ 形 2 アクティブ 形 1 ノー 形 1 マンスリー 形 1 イン 前 1 オンリー 副 1	グローバル 形 24 デジタル 形 17 スーパー 形 12 アップ 副 12 ベスト 形 5 ジュニア 形 4 オール 形 3 オンライン 形 3 パーチャル 形 3 モバイル 形 3 イン 前 3 アクティブ 形 2 キッズ 形 2 ザ 冠 1 クローズド 形 1 ハンディー 形 1 フリー 形 1 マイ 所 1 アゲイン 副 1 オンリー 副 1	グローバル 形 18 デジタル 形 17 ジュニア 形 8 アップ 副 8 スーパー 形 7 ニュー 形 7 オンライン 形 5 スペシャル 形 3 モバイル 形 3 オール 形 2 キッズ 形 2 ノー 形 2 パーチャル 形 2 イン 前 2 フォーエバー 副 2 アクティブ 形 1 ビッグ 形 1 ベスト 形 1 ウエルカム 動 1

表 3: Web での英語表現の集計

前置詞 (34 単語, 総数 16352)		all	50	* ほかいろいろ	Writtenby	28	* [人名]	
in	4852	キャンペーン*渋谷	more	47	* レッドハットニュース	producedby	27	* [組織名]
by	2504	現地レポート* [人名]	free	45	* 素材	photoby	25	* [組織名]
vs	1894	[人名] * [人名]	public	42	* 参加	welcometo	24	* 情報科学部
de	1290	メール* ゲット!	next	41	個展* 開催	Poweredby	23	* [人名]
in	1276	ジャストリオ* 神戸	Open	40	* フォーラム	Backto	22	* 広島大HP
vs	888	山形市民* 天童市民	Weekly	39	* レポート	It's	18	* オススメ!!
with	533	[人名] * オケストラ	active	35	* 手段	postedby	17	* [人名]
for	461	データ* ピアノ・レッスン	close	34	全て* です	foryou	15	メッセージ*
to	360	固定電話* 固定電話	good	34	* ウェイキ	presentedby	14	* [組織名]
from	354	文部科学省の働き* [人名]	Dear	33	* [人名]	Textby	12	* [人名]
by	269	医療施設* [組織名]	Special	29	* キャンペーン	what's	12	* ろうきん?
at	254	忘年会* 難波	Digital	27	* 出力	What's	12	* ろうきん
vs.	254	湘南* 新潟	cool	27	* ブランド	That's	11	* インタビュー
de	240	ねっと* しょっぱ	official	26	* サイト	howto	11	* 物
on	189	興津* メディア	Personal	25	* モデル	副詞 (7 単語, 総数 593)		
From	145	* こうち	Cool	24	* サイト	up	208	9,000 円*
of	77	編集部* 速報ニュース	Jr.	24	* 選手権	off	131	10 パーセント*
By	75	* [人名]	Back	23	* モデル	not	95	* オススメ
to	60	ドア* ドア	Next	23	* シーズン	only	63	土日*
For	44	* ビギナーズ	virtual	23	* 体験	all	50	同業者*
To	42	* 軽井沢	weekly	22	* パッケージ	off	30	500 円*
with	42	[人名] * [人名]	Happy	21	* 懸賞らいい!	only	16	英語*
By	27	* 監督	digital	20	* 回路	冠詞 (3 単語, 総数 364)		
About	26	* 文学部	mobile	20	* 端末	The	201	* 沖縄
about	25	* 海洋堂	special	20	* 企画	the	115	* 古書
for	25	コンサート* [組織名]	Best	19	* 記録	the	48	* 日本列島
on	24	アンケート集計* ネット	happy	18	* ハート	所有代名詞 (4 単語, 総数 357)		
Vs.	21	* [組織名]	Mobile	17	* 端末	My	181	* コレクション
since	20	* 平成 11 年 9 月 27 日	closed	16	* 会議	My	77	* ページ
at	20	コンテスト* [地名]	Jr.	16	* フルーツ教室	my	69	* テーマ
With	17	* [人名]	Virtual	15	* ゼミ訪問	my	30	* ページ
from	17	独り言* [人名]	inline	15	* 画像	接尾辞 (2 単語, 総数 202)		
by.	14	* [人名]	personal	15	* ホームページサーバ	ing	148	俳句 @ing
In	13	* [地名]	Good	14	* リフォーム	ing	54	職* 全国版求人情報
接続詞 (4 単語, 総数 3458)		Hot	14	* リスト	動詞 (9 単語, 総数 246)			
or	2236	交換* 差し上げます	Monthly	14	* セミナー	start	56	* インターネット
and	641	[人名] * [人名]	short	14	* ブーツ	Welcome	48	* えひめ
or	541	おかゆ* ご飯	Big	13	* プレゼント	Do	29	* 中国茶
and	40	パスタ* ワイン	Easy	13	* 登録	welcome	28	* [組織名] 病院
形容詞 (59 単語, 総数 3066)		Global	13	* 企業	Get	25	* 懸賞	
new	582	* おまかせプラン	offline	13	* 状態	love	22	* 熊本電波
New	399	* ぐるりん阿蘇	Hyper	12	* 和紙	enjoy	16	* 講座
main	158	* 作品	super	12	* 特急	We l c o m e	12	* [組織名] ホームページ
open	133	* 会議	Daily	11	* ランキング	Love	10	* みやざき
online	130	* 雑誌	High	11	* レベル	疑問代名詞 (1 単語, 総数 10)		
good	111	* 性能バランス	句 (23 単語, 総数 909)		What	10	* 共栄会	
New	74	* ステッカー	Welcome to	294	* 腕時計資料室			
No	72	* 残業 DAY	What's	114	* けいば			
Super	72	* アモルファス	Let's	75	* スタート!			
n e w	72	* アルバム	Presented by	34	* 関東バス			
Online	67	* 作文教室	written by	32	* [人名]			
back	65	* ナンバー	Let's	31	* テニス!			
Free	52	* 座談室	Howto	29	* アクセス			
			Produced by	29	* [組織名]			

表 4: Web での片仮名表現の集計

副詞 (5 単語, 総数 23258)		ニュー	481	* 軽量骨材	オブ	24	良品* 京都	
アップ	19615	何点*	オフィシャル	384	* 応援歌	トゥ	23	フロム市長* 職員
オフ	3373	日本語入力*	インライン	343	* 入力	動詞 (3 単語, 総数 741)		
オンリー	243	読み物*	パーソナル	287	* 感覚	エンジョイ	611	* ふくい
アゲイン	26	長野*	オール	269	* 専任教師	ウェルカム	73	* へいあん
オンリィ	1	[人名] *	ショート	269	* 丈	ウェルカム	57	* 東久留米
形容詞 (40 単語, 総数 64867)		クロス	264	* 状態	所有代名詞 (1 単語, 総数 270)			
デジタル	15319	* 処理	グッド	260	* くんまの新鮮野菜	マイ	270	* 切手
オンライン	12223	* 写真集	パブリック	236	* 空間	冠詞 (1 単語, 総数 178)		
メイン	7982	* 曲選出	ハイパー	186	* 調査隊	ザ	178	なるほど* 信州
フリー	3953	* 素材	ノー	178	* 試験	接続詞 (1 単語, 総数 132)		
ベスト	3543	* 盤	デイリー	160	* 情報	アンド	132	計 12 枚* 袋チヨコ 1
モバイル	3230	* 端末	マンスリー	91	* 壁紙	句 (3 単語, 総数 83)		
デジタル	2241	* 画像	ネクスト	81	* 牌	レッツ	73	* 不動産
スペシャル	1677	* 料金	ハンディー	73	* 機器	ザッツ	8	* 下町考
グローバル	1636	* 企業	クローズド	57	* 期間	ウェルカムトゥ	2	* お台場ラジオ会館
ホット	1578	* 情報	イージー	30	* 操作	接尾語 (1 単語, 総数 28)		
アクティブ	1442	* 商社	エクストラ	21	* 告知ページ	イング	28	遊*
スーパー	1395	* 不況	ウィークリー	20	* 検索			
ジュニア	1247	* 選手権	スーパ	19	* 熱分析システム			
オンライン	769	* 質問フォーム	アクティヴ	10	* 発想			
クール	754	* 解説	前置詞 (6 単語, 総数 1816)					
ハンディ	574	* 機	イン	1214	探検隊* [地名]			
バーチャル	569	* 博物館	デ	360	ねっと* わーく			
ビッグ	517	* 当選	アバウト	129	* 八日市			
ハッピー	499	* [組織名]	フロム	66	* 沖縄推進機構			