

Webにおける現場指示表現と、情報の画像依存度

長澤陽子（駒沢女子大学）・清水由美子（武蔵工業大学）

1 はじめに

HTML のソースのパターンから画像の位置や画像の使用頻度などは分かるが、文章の情報が画像の情報に依存するかどうかを判断するのは難しい。本稿では画像を指し示す現場指示表現に着目し、現場指示表現となりうる文字列の頻度が情報の画像依存性を反映しうるかについて調査する。ただしここでいう現場指示表現には、指示詞の現場指示的用法以外にも、時空から切り離した場合に意味をなさなくなる体言を含める。

2. 画像を指し示す現場指示表現

2.1 指示詞の現場指示的用法

指示詞の現場指示的用法とは堀口（1978）によれば「基本的には、対話・講演など話し手と聞き手が同一の空間を共有する場面において、多くの場合身ぶり・手ぶり・表情などの表現行為を伴いつつ、話し手が現に知覚していて聞き手にも知覚されるはずだとする事物を対象として、コ・ソ・ア系の語を用いて指示する用法」である。林の「代名詞が指すもの、その指し方」（1983）には、小説『夢十夜』にあらわれた現場指示的用法の例が38例あるが、そのうち30例は引用表現の中で用いられており、7例は主人公の心のつぶやきである。この結果は指示詞の現場指示的用法が純粋な書き言葉の中にあらわれにくいことを示唆している。

註：残りの1例は慣用句である。本稿では慣用句は現場指示的用法ではないとする。

一方 Web の文章には、本来談話で用いられるはずの現場指示的用法が多くあらわれる。#1、#2 は画像そのものを、#3 は画像内の事物を指し示している。#4、#5 は文字列「ここ」自体を指す。#4、#5 の扱いは難しいが、Web という現場から切り離すと意味をなさなくなるため、現場指示的用法と解釈した。なお、#6、#7 はメタ的用法であり、現場指示的用法ではない。

註：特に断らない限り、#で示すものは、2月22日 goo により2コンテンツ以上ヒットしたパターンである。

#1 この写真は…。

#4 詳細はここ

#2 これは…の写真です。

#5 ここをクリックしてください。

#3 この建物は…である。

*#6 このページは…。

*#7 ここは…のページです。

2.2 画像を意味する体言

コ系指示詞以外にも現場指示的に用いられ画像を指し示す語は多い。その一つに画像を意味する語（「写真」「絵」「イラスト」「図」「地図」「表」「グラフ」）がある。この現場指示表現（註）は紙の媒体においても用いられる。

#9 写真は…の時のものです。

#10 …図参照。

註：指示詞が文章中で図・表・絵などに言及している場合に、これを「文脈指示」とする方法もある。

しかし、ここでは、“文章以外のもの（画像など）内の文脈”という難しい問題に立ち入ることを避けるために、「擬似的な時空共有の場における現場指示」ととらえる立場をとる。

2.3 画像内の事物を意味する体言

#11 顔をクリックしてください。

#11 の「顔」は、時空により特定されており、そこから切り離すと意味をなさない。画像内の事物を意味する体言として多用される語は、ジャンル（テーマ）によって異なる。

2.4 位置を意味する体言

#12 下は…の写真である。

#14 中央の建物は、…で

#13 左の写真は、…で

#15 左側にあるのは…で

#12、#13 は、文章に対する画像の位置を表すケースで、#14、#15 は画像内の事物の位置を表すケースである。

3. 非画像依存情報における指標出現頻度

「この」「これ」「写真」「上」「下」「左」「右」「中央」の文字列を指標とし、情報が画像に依存しない 16 作品における頻度を調査し表にまとめた。これは、画像依存情報と非画像依存情報との間に文字列出現状況の差があるかどうかをみる（4 節）ためである。対象とするサンプルは、1500 字以上の長さを持ち、クリッグせずに見ることのできる範囲内に写真を含むものをとった。偏りをさけるため、1998 年 2 月現在 Yahoo Japan のトップメニューの項目それぞれからパソコンを用いてランダムに選択したが、「各種資料と情報源」においては、20 試行中該当するものにあらず作業者が選んだ。

頻度として、文字列の出現度数／文数、文字列の出現度数／文字数 * 100 の 2 種類を算出した。サンプル数が少ないため、統計による処理は行っていない。

その結果、値は表 1 のように分布した。文数に対する「この」「これ」の頻度はほぼ正規分布、位置を表す語の頻度はやや右に裾をひいた形である。「写真」の出現頻度は、文章のジャンルに強く左右されている。文字列に対する頻度の分布は二標本問題を暗示する形になっている。

註：二重線は非画像依存情報 16 例の上限であり、アルファベットは、次節で言及する画像依存情報の値を示している。

表 1：指標別にみた、出現頻度の分布

指標の出現 度数／文数	「この」 「これ」	「写真」	「上」「下」「左」 「右」「中央」	指標の出現度数 ／文字数 * 100	「この」 「これ」	「写真」	「上」「下」「左」 「右」「中央」
0~0.05	2	B 13	0	0~0.05	2	B 13	0
0.05~0.1	4	A 1	2	0.05~0.1	1	0	0
0.1~0.15	6	1	7	0.1~0.15	3	2	2
0.15~0.2	3	0	4	0.15~0.2	2	0	1
0.2~0.25	1	7	A 2	0.2~0.25	2	A 0	4
0.25~0.3	0	0	1	0.25~0.3	4	0	6
0.3~0.35	AB	0	0	0.3~0.35	0	0	0
0.35~	0	0	0	0.35~0.4	2	0	1
	0	0	B 0	0.4~0.45	0	0	0
	0	0	0	0.45~0.5	0	0	0
	0	0	0	0.5~0.55	0	7	2
	0	0	0	0.55~	0	0	0
	0	0	0		AB 0	0	AB 0

4. 画像依存例における指標出現頻度

4.1 コンテンツ A

次に画像依存例における指標出現頻度を、上述の16例と比較する。

コンテンツ A には、画像に依存する情報（旅行記）のページと、画像に依存しない情報（旅行記ではあるが建造物の解説が中心）のページとが存在する。両者とも、上が写真、下が文章というデザインである。画像依存箇所と非画像依存箇所の値を比較したものが表2である。

表2：文字列の出現頻度（コンテンツ A）文数・文字数は、表3参照

	指標数/文数			指標数/文字数*100		
	「この」「これ」	「写真」	「上」「下」「左」「右」「中央」	「この」「これ」	「写」「真」	「上」「下」「左」「右」「中央」
画像依存箇所	0.347	0.059	0.247	1.232	0.209	0.877
非画像依存箇所	0.072	0.072	0.155	0.188	0.188	0.403

画像依存箇所における文字列「この」「これ」の頻度は、3節の16作品の範囲を逸脱している。文字列「上」「下」「左」「右」「中央」も同様に特異な値をとっている。しかし、非画像依存箇所は、非画像依存作品16例の範囲内である。

表3：指示詞の出現度数（コンテンツ A）

	文数	文字数	この	これ	その	それ	あの	あれ
画像依存箇所	170	4788	画像を指示する指示詞	22	29	0	0	0
			画像を指示しない指示詞	4	4	12	5	1
非画像依存箇所	97	3718	画像を指示しない指示詞	5	2	10	7	0

画像を指示するコ系指示詞が多く、画像依存箇所と非画像依存箇所とではコ系指示詞が出現する頻度の差（コ系指示詞を含む文と他の文との割合の差によった）が有意（ χ^2 検定・水準5%・イエーツの補正）になっている。このことが文字列「この」「これ」の頻度に影響している。

4.2 コンテンツ B

表4：文字列の出現頻度（コンテンツ B）文数・文字数は、表5参照

	指標数/文数			指標数/文字数*100		
	「この」「これ」	「写真」	「上」「下」「左」「右」「中央」	「この」「これ」	「写」「真」	「上」「下」「左」「右」「中央」
画像依存箇所	0.300	0	0.400	0.657	0	0.876
非画像依存箇所	0.361	0	0.062	0.687	0	0.118

コンテンツ B は考古学研究室のページであり、出土品や発掘現場の写真が掲載されている。コンテンツ A と同様、画像依存箇所の指標頻度が高いが、非画像依存箇所におけるコ系指示詞の頻度も同様に高い（表4）。

表5：指示詞の出現度数（コンテンツ B）

	文数	文字数	コ現場	コ現場以外	ソ現場	ソ現場以外	ア現場	ア現場以外
画像依存箇所	70	3197	16	5	1	11	0	0
非画像依存箇所	96	5092	0	35	0	10	0	0

この作品の非画像依存箇所では現場指示的用法以外のコ系指示詞（ことに文脈指示的用法）が多用されている。画像によって説明されていた部分が文章で説明されており、文と文をつなぐ文脈

指示的用法が多く用いられている。形態を同じくする別の語（もしくは用法）の多用という問題は、

(1) 形態素解析により解決可能な場合…「下痢」「これこれしかじか」「たこの足」「写真集」

(2) 用法を確定する必要のある場合…道案内における「右に曲がって」

もある。このような例が多用される場合には、文字列を使用しても精度の高い判断は望めない。

4.3 コンテンツC

コンテンツCには、将棋の駒の書体についての説明が書かれている。画像依存情報であるにも関わらず、頻度の値は非画像依存情報の範囲内である。

表6：文字列の出現頻度（コンテンツC）文数・文字数は、表7参照

	指標数/文数			指標数/文字数*100		
	「この」「これ」	「写真」	「上」「下」「左」「右」「中央」	「この」「これ」	「写真」	「上」「下」「左」「右」「中央」
C (画像依存)	0.143	0	0.190	0.265	0	0.353

表7：指示詞の数（コンテンツC）

	文数	文字数	コ現場	コ現場以外	ソ現場	ソ現場以外	ア現場	ア現場以外
C (画像依存)	21	1134	2	1	0	2	0	0

コンテンツCは、画像と文章とが対応した構成（上が「王将」の駒のイラスト、下が説明）になっており、しかも話題は駒の書体に限定されている。したがって、コ系指示詞や位置を示す体言を用いる必要はなく、このことが文字列の頻度の低さにつながったと考えられる。

5. おわりに

画像依存情報では現場指示表現の多用がめだつため、今回指標としたような文字列の頻度が画像依存性を反映する場合がある（4.1、4.2節）が、これは全てのページに該当するわけではない。

「写真」の頻度（3節参照）に明らかのように、現場指示的体言の頻度はジャンル・テーマと密接に関わる。また、4.2節の例は、別の用法が多用された場合の危険性を示しており、4.3節の例は、現場指示表現のほかにも画像と文章を対応づける方法があることを示している。

情報化が進んだ社会では、文章と画像との関係は興味深い問題となる。しかし、文章と画像との関係を機械的に指摘するためには以下の作業をする必要がある。

1. 現場指示表現とジャンル・テーマの関係を分析する。
2. 画像依存情報の語彙を調査し、語相互の共起関係による意味の絞り込みを検討する。
3. 画像依存情報の文章の形態や、用いられているHTML技術の傾向を分析する。

★参考文献

1. 堀口和吉(1978)「指示語『コ・ソ・ア』考」『論集日本文学・日本語5』角川書店
2. 林四郎(1983)「代名詞が指すもの、その指し方」『朝倉日本語新講座 運用I』朝倉書店
3. 馬場俊臣(1989)「要約文の指示語使用の特徴」『文章構造と要約文の諸相』くろしお出版
4. R.E. Horn (1991)『ハイパーテキスト情報整理学』日経BP
5. 金水敏・田窪行則(1992)「談話管理論からみた日本語の指示詞」『認知科学の発展 Vol3』講談社サイエンティフィック
6. 金水敏・田窪行則(1992)『日本語研究資料集・指示詞』ひつじ書房
7. 東京大学教養学部統計学教室編(1996)「統計学入門」東京大学出版会