

독일어 광고슬로건의 언어적 특성들에 대한 연구*

- 코퍼스언어학적인 접근

이민행 (연세대)

1. 서론

이 연구에서는 독일어 광고슬로건 코퍼스를 활용하는 방안과 더불어 슬로건들의 다양한 언어적 특성과 시대별 사용양상에 대해 논의한다.

지금까지 독일어 광고슬로건에 대한 많은 연구들이 있었으나 광고슬로건 코퍼스를 대상으로 본격적으로 언어적 특성들을 밝혀낸 연구는 없었다.¹⁾

광고슬로건을 대상으로 한 연구들 중에서 잘 알려진 독일어권 연구는 Römer (1976)와 Baumgart (1992)의 연구이다. 전자의 경우, 광고슬로건에 나타난 수사학적인 문체 Figuren를 중심으로 연구를 수행하였고 후자의 경우, 1960년대와 1980년대에 사용된 핵심어휘들을 비교하여 시대상의 차이를 밝혀낸 바가 있다. 국내 독어학계에서는 김원식 (2001)에서 광고슬로건의 기능에 대한 연구가 이루어졌다.²⁾

본 연구는 필자가 직접 구축한 코퍼스 GLOS를 토대로 하여 광고슬로건의 언어적 특성들을 밝혀내는 것을 연구목적으로 한다.³⁾ GLOS 코퍼스는 어휘 규모가 토큰을 기준으로 90,567개 (15,757 문장)에 이르며 1,770년부터 2,005년까지 독일어권의 광고에 나타난 슬로건으로 구성되어 있다.⁴⁾ 이 코퍼스는 검

* 이 논문은 2013년 정부 (교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (2013S1A6A4016585).

- 1) Hennemann (2013)에서는 코퍼스를 기반으로 한 양적, 질적 연구방법들이 논의된다.
- 2) 오장근 (1999)에서는 슬로건이 광고기호의 해석과 관련하여 어떤 기능을 수행하는지를 논의하고, 정동규 (2000)는 슬로건에 나타난 단어유회의 양상을 탐구한다.
- 3) 코퍼스 GLOS (German Language Of Slogans)의 토대가 된 언어자료는 광고슬로건 사이트 www.slogans.de로부터 추출한 것이다. 슬로건들을 수집하는 과정에서 연세대 최경은 박사의 도움을 받았다.
- 4) 1770년의 광고 슬로건은 “Aus Freude am Tabak.”이 유일하다.

2. 독어학 제29집

색엔진 CQP에 의해 검색이 가능하도록 CWB-시스템의 하나로 구축되어 있다.⁵⁾

코퍼스 GLOS에는 어휘형태 word, 품사 pos 및 레마 lemma에 대한 정보 외에 문장단위 (s) 정보 및 언어 (text_lang), 연도 (text_year), 산업 (text_industry) 및 브랜드 (text_brand)에 대한 정보가 포함되어 있다.

2. 본 론

이미 언급한 바 있듯이 코퍼스 GLOS에는 연도에 대한 정보가 포함되어 있다. 이 정보를 활용하여 시대별로, 품사별로 어떤 어휘들이 많이 사용되어 있는지를 연구할 수 있다. 언어가 사회문화를 반영한다는 입장을 지지하는 어휘 데이터를 찾아낼 수 있다면 코퍼스가 유용성을 갖는다고 말할 수 있을 것이다. 다음 예를 살펴보자.

```
(1) GLOS; T=[pos="NN"]:.match.text_year="195.*"; count by lemma >
"freqNN1950s.txt";
```

위 (1)은 세 가지 절차를 일괄명령으로 묶어놓은 것이다. 첫째 절차는 코퍼스 GLOS를 불러내라는 명령이고, 둘째 절차는 'text_year="195.*"'라는 정보를 포함하고 있는 문장들에 한정하여 일반명사 (NN)를 검색하라는 지시이며, 셋째 절차는 검색결과를 'lemma'를 기준으로 하여 출현빈도와 함께 'freqNN1950s.txt'라는 파일에 저장하라는 의미이다. 저장된 파일의 내용을 일부만 보이면 다음과 같다.

(2) 빈도	키워드
18	Frau
18	Welt

5) CQP (corpus query processor)와 CWB (corpus workbench)에 대해서는 Evert (2010), Hardie (2012) 및 이민행 (2012) 참조.

12	Haut
11	Herz
10	Haar
10	Liebe
10	Pause
10	Sicherheit
10	Strumpf
10	Tag
10	Zeit
10	Wäsche

이 데이터는 1950년대 사용 명사의 전체 목록 중에서 빈도순위 1위-12위까지만 추출한 것이다. 위 (1)의 둘째 절차와 셋째 절차를 아래 (3)과 같이 수정할 경우에 1960년대의 데이터를 추출할 수 있다.

(3) A=[pos="NN"] ::match.text_id="196.*" ; count by lemma > "freqN1960s.txt";

이러한 과정을 반복하면 1950년부터 2000년대 까지 사용된 명사에 대한 데이터를 모두 모을 수 있다. 이 가운데서 빈도수 1위부터 20위까지만을 표로 정리하면 [표 1] 및 [표 2]와 같다.

순위	1950년대		1960년대		1970년대	
	빈도	명사	빈도	명사	빈도	명사
1	18	Frau	54	Welt	31	Welt
2	18	Welt	26	Mann	29	Geschmack
3	12	Haut	23	Hand	22	Haut
4	11	Herz	19	Haar	21	Qualität
5	10	Haar	18	Geschmack	20	Tag
6	10	Liebe	18	Tag	19	Mann

4 독어학 제29집

7	10	Pause	18	Qualität	18	Frische
8	10	Sicherheit	17	Haut	16	Haar
9	10	Strumpf	17	Kind	16	Leben
10	10	Tag	17	Zeit	16	Mensch
11	10	Zeit	16	Kraft	16	Sicherheit
12	10	Wäsche	15	Wäsche	13	Haus
13	8	Freiheit	14	Fortschritt	13	Natur
14	8	Freude	14	Mensch	12	Spaß
15	8	Seife	14	Sekt	12	Zeit
16	8	Weinbrand	13	Frische	12	Schönheit
17	8	Qualität	13	Genuss	11	Freude
18	7	Bier	13	Leben	11	Gesundheit
19	7	Vertrauen	12	Bier	11	Herz
20	7	Glück	12	Kaffee	11	Kraft
					11	Vernunft

<표 1> 1950년대-1970년대 명사의 사용양상

순위	1980년대		1990년대		2000년대	
	빈도	명사	빈도	명사	빈도	명사
1	61	Geschmack	43	Leben	186	Leben
2	46	Haut	27	Idee	106	Welt
3	32	Haar	23	Welt	88	life
4	32	Leben	19	Spaß	81	Idee
5	32	Qualität	17	Haut	77	world
6	30	Welt	15	Mensch	72	Mensch
7	29	Natur	14	Mann	70	Zukunft

8	27	Idee	14	Sicherheit	68	Zeit
9	23	Duft	14	Zukunft	64	Natur
10	21	Pflege	12	Natur	54	Haus
11	20	Mode	12	Zeit	50	Erfolg
12	19	Genuss	11	Genuss	46	Genuss
13	18	Schönheit	11	Tag	46	Haut
14	18	Stück	11	Technik	44	Energie
15	17	Kraft	10	world	43	business
16	17	Technik	9	Energie	43	Qualität
17	17	Zukunft	9	Mode	41	Geschmack
18	16	Sicherheit	9	life	40	Beste
19	15	Frische	9	Qualität	40	Gesundheit
20	15	Tag	8	Geld	39	Kraft
	15	Weg	8	Internet		
			8	Gefühl		

<표 2> 1980년대-2000년대 명사의 사용양상

시대별 구분을 하지 않고 1950년대부터 2000년대까지 통틀어서 어떤 명사들이 광고슬로건에 빈번히 나타났는지를 확인하고자 하면 검색식을 다음과 같이 단순화하면 된다.

(4) A=[pos="NN"] ; count by lemma > "freqN.txt";

검색결과중에서 1위부터 20위까지만 제시하면 다음 (5)과 같다.

(5) 빈도 키워드
 295 Leben
 263 Welt
 163 Geschmack
 163 Haut

6 독어학 제29집

145	Idee
134	Mensch
133	Natur
133	Qualität
131	Zeit
114	Zukunft
113	Tag
106	Haus
101	Haar
100	Genuss
100	Mann
97	Sicherheit
97	life
94	Kraft
94	world
81	Spaß

이런 명사들은 시대를 초월해서 높이 평가되는 보편적인 가치를 담고 핵심어들이라고 할 수 있다 (Baumgart 1992: 123).

동일한 방법론을 이용하여 형용사의 시대별 핵심어들도 추출할 수가 있다. 시대를 초월하여 빈번히 출현하는 형용사들의 목록 1위부터 20위까지 제시하면 다음 (6)과 같다.

(6) 1950년대부터 2000년대까지 사용된 형용사 목록

빈도	형용사
460	gut
234	neu
142	groß
132	schön
90	ganz
87	gesund
69	klein

65	fein
59	stark
56	erst
56	voll
52	natürlich
51	frisch
50	modern
48	rein
40	richtig
38	echt
35	deutsch
33	jung
32	@ord@
29	hoch

이 목록에서 확인할 수 있듯이 높은 가치를 표현하는 형용사들은 시대를 뛰어넘어 사랑을 받는다. 이들 중 대표적인 것은 gut, schön, neu, groß, einfach 등이다.

광고 코퍼스를 이용하여 파생어에 관한 연구도 수행할 수 있다. 이를 테면 파생접미사 -bar로 끝나는 형용사들이 광고에 얼마나 나타나는지를 확인하기 다음과 같은 검색식을 사용할 수 있다.

(7) T = [pos="ADJ.*" & lemma="*.bar"]; count by lemma > "freqADJbar.txt";

이 검색식을 실행하면 다음 (8)과 같은 결과를 얻게 된다.

(8) 빈도	키워드
18	wunderbar
13	kostbar
10	sichtbar
7	spürbar
5	unverkennbar

8 독어학 제29집

3	unsichtbar
2	unbezahlbar
2	unverwechselbar
1	bezahlbar
1	haltbar
1	meßbar
1	planbar
1	steuerbar
1	unberechenbar
1	unschlagbar
1	unvergleichbar
1	unverzichtbar
1	unzerreißbar
1	wanderbar
1	hörbar
1	wählbar

위 검색식 (7)의 위치속성 lemma의 값 안에 위치한 ‘bar’ 대신에 다른 파생 접미사 ‘wert’를 대체해서 실행을 시키면 다음 (9)와 같은 결과를 얻는다.

(9) 빈도	키워드
26	preiswert
8	unbeschwert
3	lebenswert
3	liebenswert
2	begehrenswert
2	bewundernswert
1	beneidenswert
1	empfehlenswert
1	preiwert

두 가지 결과를 비교해 볼 때 흥미로운 것은 광고와 관련되어 있어 파생접

미사 ‘-wert’가 접미사 ‘-bar’보다 많이 활용될 것 같은데도, 검색결과는 그 반대라는 점이다.

형용사와 명사가 인접해서 나타나는 용례들도 코퍼스 검색을 통해 추출할 수 있다. 이를 위해 다음과 검색식을 사용하면 된다.

(10) T=[pos=“ADJA”][pos=“NN”]; count by lemma > “freqAdjNN.txt”;

다음은 검색결과로부터 1위부터 10위까지만 선택한 것이다.

(11) 빈도	키워드
19	gut Idee
18	gut Geschmack
17	ganz Tag
17	ganz Welt
16	gut Laune
14	voll Geschmack
13	schön Haar
12	fein Unterschied
12	gut Hand
12	gut Gefühl

이 목록을 통해서 대략 어떤 명사에 어떤 형용사가 어울리는 지를 확인해 볼 수 있다. 광고의 특성상, 형용사들은 핵어 명사의 원형적인 속성을 가장 잘 드러내주는 어휘들이다.

인접하는 부사와 형용사간에도 긴밀한 관계가 성립할 수 있는데, 이 데이터를 추출하기 위한 검색식은 다음 (12)와 같다.

(12) T=[pos=“ADV”][pos=“ADJ.*”]; count by lemma > “freqAdvAdj.txt”;

검색결과로부터 출현빈도가 4이상인 예들만 선택한 것이 다음 (13)에 제시된다.

10 독어학 제29집

(13) 빈도	검색어
40	so gut
20	immer gut
16	so einfach
10	ganz schön
10	noch gut
10	so leicht
10	so nah
9	so frisch
8	ganz einfach
8	so schnell
7	so wichtig
6	so sauber
6	natürlich gesund
5	immer frisch
5	immer neu
5	nur echt
5	so schön

이 데이터를 살펴보면, 특정한 부사와 특정한 형용사가 서로 결합한다기보다, 대체적으로 정도를 표현하는 부사들이 높은 가치를 표현하는 형용사들과 어울린다고 해석하는 편이 옳을 것 같다. 이런 맥락에서 부사와 형용사가 언어관계를 이룬다는 주장을 하기에 데이터가 너무 빈약하다고 할 수 있다.

광고에서 사용되는 구문의 특성에 대해서 살펴보는 것도 흥미로운 작업인데, 이목을 끄는 구문중 하나가 결과구문이다.⁶⁾ 동사 **machen**이 이끄는 결과구문을 검색하기 위해 다음과 같은 검색식을 사용한다.

(14) T= [lemma="machen"][] {0,1} [pos="NN"] [pos="ADJ.*"]; cat T > "conc MachenSatz.txt";

6) 코퍼스를 이용한 결과구문의 연구에 대해서는 이민행 (2012: 356) 참조.

검색결과는 64개 용례가 추출되었는데, 모두가 광고제품을 사용함으로써 목표수용자가 어떤 편익을 얻게 된다는 메시지를 전하고 있다. 다음에 몇 가지 예가 제시되어 있다.

(15)

- 550: Diplona <macht das Haar gesund>.
- 779: <Macht jedes Teil schnell> wieder heil!
- 1000: <Macht Ihr Haar schmiegsam> bis in die Spitzen.
- 2985: Miele <macht's der Hausfrau leichter>.
- 4490: <Macht die Haut gesund> und schön.
- 6091: Bosch <macht Frauenwünsche wahr>.
- 6876: <Macht Ihre Zähne strahlend> weiß.
- 8977: Haribo <macht Kinder froh> und Erwachsene ebenso.
- 12998: SC <macht Männer mutig>.
- 18596: <Macht den Kreislauf fit>!
- 22519: Europcar <macht das Automieten leichter>.
- 25234: Bertelsmann <macht Freizeit schön>!
- 30444: Zentis <macht das Leben fruchtiger>!

결과구문의 검색과 관련하여 이 구문에 나타난 목적어 명사와 보어 형용사 간의 상관관계도 살펴볼 수 있는데, 이를 위해서 다음과 같은 검색식을 사용한다.

(16) T= [lemma="machen"]{0,1}@[pos="NN"][[pos="ADJ.*"]; group T matchend
lemma by target lemma > "freqMachenNnAdj.txt";

이 검색식은 group 함수와 준거점 연산자 target를 이용하여 빈도를 추출하는 작업을 수행한다. 그 결과의 일부는 다음 (17)과 같다.

12 독어학 제29집

(17)

Mensch	weltweit	5
Leben	lang	5
Macht	einfach	4
Leben	schön	4
Haut	gesund	4
Welt	groß	4
Idee	gut	4
Welt	voll	4
Greifen	nah	3
Zeichen	gut	3
Küche	kalt	3
Kind	lieb	3
	froh	3
Leben	leicht	3
Gold	wert	3
Tag	gut	3

이 단어쌍들은 앞서 제시한 용례들에 나타나 있는 것들이 대부분이어서 어떤 맥락에서 사용되는지를 짐작하기가 어렵지 않다.

마지막으로 평서문, 명령문 및 의문문 등 문장유형의 분포는 어떻게 되는지 살펴보자. 이를 위해 다음과 같은 검색식을 사용한다.

(18) T=[word="^.\{?\}"]; count by word > "freqSatzTypen.txt";

이 검색식을 실행한 결과 얻은 데이터는 다음과 같다.

(19) 빈도	문장유형
15358	.
1344	!
130	?

이 데이터는 평서문이 가장 많이 쓰인다는 사실을 확인시켜준다. 이 맥락에서 관심을 가져볼 만한 이슈는 의문문은 어떤 용도로 사용되는 지를 알아보는 것이다. 이를 위해 검색식 (18)을 변형하여 다음과 같은 검색식을 사용하면 용례를 추출할 수 있다.

(20) T=[word="{?}"]; cat T > "concFrageSatz.txt";

이 검색식을 실행한 결과 가운데서 몇 가지만을 제시하면 아래와 같다.

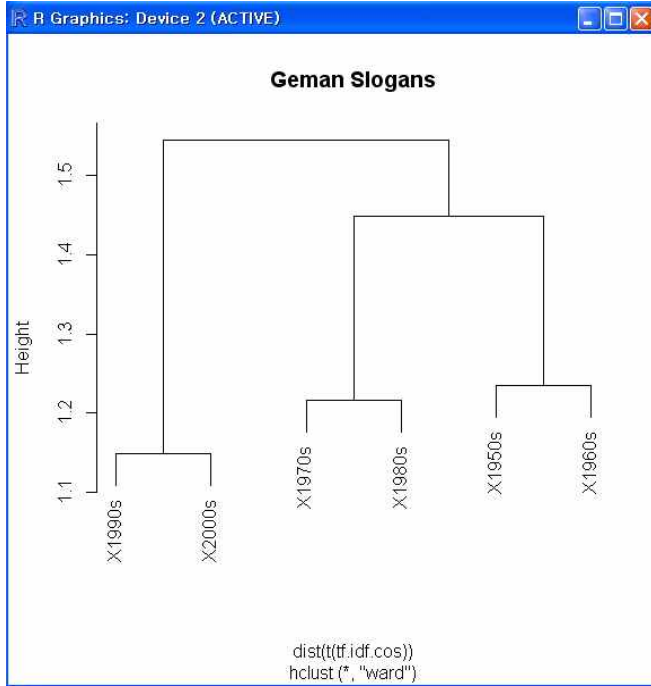
- (21)
- 1413: PEZen Sie schon<?>
 - 4956: Kennen Sie schon Calgon<?>
 - 15867: Was trinken wir<?> Schultheiss Bier.
 - 21208: Auch schon probiert<?>
 - 31660: Ist der neu<?> Nein, mit Perwoll gewaschen.
 - 35815: Wo ist der Deinhard<?>
 - 38446: Was würdest du tun<?>
 - 39411: Alles klar<?> Alles Hansaplast!
 - 40358: HB<?> Ich auch.
 - 45220: Alles Müller oder was<?>
 - 50979: Ist das nur ein Automobil<?>

의문문의 형식을 취한 슬로건들을 면밀히 검토해 보면 세 가지 특성을 발견하게 된다. 첫째는 소위 답을 이미 알고 질문을 하는 수사어문문이 많이 쓰인다는 사실이다. 둘째는 첫째 특성과 관련되기도 하는데, 초점첨사 schon이 의문문에 많이 나타난다는 점이다. 셋째 특성은 대화를 시작할 목적으로 목표 수용자에게 말을 거는 용법이다.

각 시대별 상위빈도 명사들을 바탕으로 R 언어를 이용하여 시대간의 유사도를 측정된 결과 아래와 같은 그래프를 얻었다.⁷⁾

7) 계층적 군집분석을 위한 R-스크립트의 작성시에 연세대 응용통계학과 김철웅 교수의 도움을 받았다. 또한 이 스크립트를 실행할 때 핵심이 되는 행렬데이터를 얻

<그림 1> 시대별 슬로건들의 유사도



그래프 상에서 ‘X1950s’은 1950년대 슬로건 문서를, ‘X1960s’는 1960년대 슬로건의 상위빈도 명사목록으로 이루어진 문서를 의미한다. 그리고 ‘X1970’은 1970년대 슬로건의 상위빈도 명사목록을 가리키며 ‘X1980s’는 1980년대 슬로건의 상위빈도 명사목록을, ‘X1990s’는 1990년대 슬로건의 상위빈도 명사목록을 그리고 ‘X2000s’는 2000년대 슬로건의 상위빈도 명사목록을 의미한다. 유사도를 구하기 위해 행렬데이터로 각 시대별 상위빈도 1위부터 50위까지를 포함시켰다.

기 위해서 빈도테이블로부터 변환한 필요가 있는데, 이 과정에 사용한 Perl 스크립트의 작성 과정에서 뮌헨대의 이영수박사가 도움을 주었다. R 프로그램을 언어학적인 연구에 활용하는 방법 전반에 대해서는 Gries (2008) 참조.

위의 그래프 상에서 기준 ‘Height’가 낮을수록 문서들 간의 유사도가 높은 것으로 해석된다. 따라서 1990년대 슬로건과 2000년대 슬로건이 가장 유사하고, 그 다음으로 1970년대 슬로건과 1980년대 슬로건이 유사도가 높다. 이어서 1950년대와 1960년대 슬로건이 서로 유사하다. 마지막으로 1950년/1960년 짝과 1970년대/1980년대 짝 간에도 유사도가 관찰된다는 분석도 설득력이 있다.

시대들 간의 유사도를 측정하여 그래프로 표상하기 위해 일반적으로 군집 분석 cluster analysis에 사용되는 R-스크립트를 이용했는데, 그 내용은 다음 (22)와 같다.⁸⁾

```
(22)
data <- read.csv("6decadeMatix.csv")           /* @1 */
tf <- data[,2:14]                               /* @2 */
idf <- rowSums(tf >= 1)                         /* @3 */
norm_vec <- function(x) {x/sqrt(sum(x^2))}     /* @4 */
tf.idf <- log2(6/idf) * tf                     /* @5 */
tf.idf.cos <- apply(tf.idf, 2, norm_vec)       /* @6 */
fit <- hclust(dist(t(tf.idf.cos)), method="ward") /* @7 */
plot(fit, main="German Slogans")               /* @8 */
```

스크립트의 첫 줄은 행렬포맷으로 정리된 데이터파일 ‘6decadeMatix.csv’을 읽어들이어서 ‘data’라는 이름을 붙이려는 명령이다.⁹⁾ 둘째 줄은 이 ‘data’ 파일의 제2열부터 제14열에 정리된 빈도정보를 추출해서 ‘tf’ (term frequency, 개별 용어빈도)라는 속성의 값으로 저장하라는 내용이다. 셋째 줄은 어휘들의 상대 빈도를 구하기 위한 함수식으로 각 행별로 기록된 모두 빈도를 합하여 ‘idf’ (inverse document frequency, 역 문서빈도) 삼으라는 명령이다.¹⁰⁾ 이어 넷째 줄은 일반적으로 비교대상 문서의 전체 규모를 고려하여 개별 어휘빈도 데이터

8) 이 스크립트의 작성을 위해 연세대 응용통계학과의 김철웅 교수의 도움을 받았다.

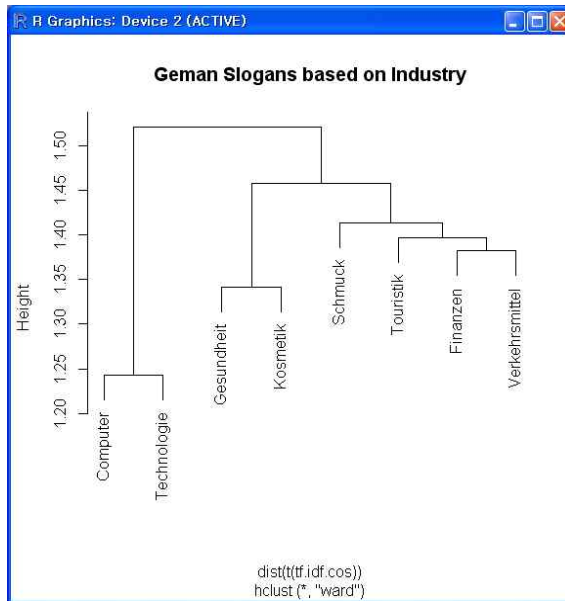
9) 행렬 형식의 파일을 준비하는 과정도 여러 단계를 거친다.

10) ‘역 문서빈도’의 개념에 대해서 스탠포드 대학의 웹사이트 (<http://nlp.stanford.edu/IR-book/html/htmledition/inverse-document-frequency-1.html>) 참조.

의 가중치를 산출하는 표준화공식을 정의한다. 다섯째 줄은 문서의 숫자와 ‘개별 용어빈도’와 ‘역 문서빈도’를 이용하여 ‘tf.idf’ 값을 산출하라는 명령이다. 이렇게 얻어진 ‘tf.idf’의 값을 넷째 줄에서 정의한 표준화공식을 적용하여 코사인 값을 구하는 과정이 여섯째 줄에 기술되어 있다. 코사인 값은 문서들 간의 거리를 나타내는 직선간들 간의 각도를 나타내므로 그 값이 작으면 작을수록 문서들 간의 유사도가 높아지는 속성이 있다. 일곱째 줄이 이 스크립트의 핵이 되는데, 여기에서 최종적으로 군집 clustering의 정도를 계산하는 작업이 수행된다. 이 스크립트에서는 ‘ward’라는 계층적 군집분석의 한 방법론으로 계층적 군집도 (함수 ‘hclust’)를 산출하여 ‘fit’라는 속성에 할당한다. 마지막 줄은 앞줄에서 얻은 군집도 ‘fit’를 그래프로 나타내는 과정을 보여주는데, 속성 ‘main’의 값으로 제시된 "German Slogans"는 그래프에 붙일 명칭을 가리킨다.

이러한 계층적 군집분석 방법론을 산업별로 분류한 슬로건들을 대상으로 유사도를 측정할 결과는 다음 그림과 같다.

<그림 2> 산업별 슬로건들의 유사도



그래프 상에서 Computer 분야와 Technologie 분야 간의 유사도가 가장 높게 나타나 있는데, 이는 상식에 잘 들어맞는 결과이다. 그 다음으로 유사도가 높은 작은 Gesundheit 산업과 Kosmetik 산업이다. 이 점도 직관에 비추어 이해가 가능하다. 나머지 네 업종들 간에도 유사도에 있어 차이가 있기도 하지만 친소 관계가 대체적으로 이해할만한 수준이다. 이 유사도 그래프를 산출하기 위해, 산업별로 명사 상위빈도 1위부터 30위까지를 기초자료로 사용했다. 그런데 30위에 있는 동일한 빈도를 가진 단어가 너무 많을 경우에는 빈도수를 한 단계 높여서 끊었다.

유사도가 가장 높은 Computer 분야와 Technologie 분야 간의 유사도 측정에 사용된 어휘 목록과 빈도는 다음 [표 3]과 같다.

순위	Computer		Technologie	
	빈도	어휘	빈도	어휘
1	11	business	8	Zukunft
2	11	power	8	world
3	9	software	7	innovation
4	8	technology	7	Lösung
5	8	work	6	future
6	7	future	6	motion
7	7	way	5	Partner
8	6	Computer	5	System
9	6	Idee	5	Technik
10	6	company	4	Idee
11	6	game	4	Welt
12	5	Erfolg	4	company

13	5	Mensch	4	life
14	5	Software	4	technology
15	5	System	3	Aluminium
16	5	mind	3	Antrieb
17	5	network	3	Erfolg
18	5	Lösung	3	Industrie
19	4	life	3	Kompetenz
20	4	place	3	Leben
21	4	web	3	Tor
22	4	world	3	Verbindung
23	3	Internet	3	Zeit
24	3	Kommunikation	3	success
25	3	Kraft	3	work
26	3	Service		
27	3	Technologie		
28	3	Welt		
29	3	Zukunft		
30	3	entertainment		
31	3	fun		

<표 3> Computer/Technologie 분야 어휘분포

위의 표를 살펴보면, 두 영역의 광고슬로건에서 영어어휘가 매우 많이 사용된다는 사실도 알 수 있다.

영어만으로 된 슬로건은 전체 15,759개의 14.5% 정도에 해당하는 2,288개이다. 아래 표는 언어사용의 분포를 보여준다.

언어	슬로건 수
독일어 (de)	13,451
영어(en)	2,288
독-외 혼용 (de_fw)	20
합 계	15,759

<표 4> 언어 분포

언어사용에 관한 분포를 얻기 위해 필요한 검색식은 다음 (23)과 같다.

```
(23) T=<s>[word=".*"]+<s>;
      group T match text_lang > "freqGlosLang.txt";
```

영어슬로건은 2000년대 들어서 급격히 증가하는 추세에 있는데, 이러한 사실은 다음과 같은 검색결과로 확인할 수 있다.

연도	어휘빈도
2004	717
2003	664
2002	222
2000	128
2001	123
2005	106
1999	58
1998	55
1997	31
1995	17
1991	16
1996	13
1988	11

1994	11
1989	10
1984	10
1990	10
1993	9
1973	6
1980	6

<표 5> 연도별 영어어휘 출현빈도

이 검색결과는 영어어휘의 사용빈도가 매년 증가한다는 사실을 보여주는데 특히 1994년과 2004년의 결과를 비교해 보면 10년 사이에 영어어휘의 사용빈도가 대략 70배 성장한 것을 확인할 수 있다. 이러한 변화는 슬로건의 수가 1994년 170개에서 2004년 3,108개로 18배 확장된 것과는 무관하지는 않으나, 이런 데이터를 감안하더라도 영어의 사용이 순수하게 5배 정도 늘어난 것이다. 위의 연도별 빈도통계는 언어와 연도간의 상호작용을 보여주는 한 단면인데, 아래 (24)와 같은 검색식을 통해 산출된다.

```
(24) T=<s>[word=".*"]+<s>:: match.text_lang="en";
      group T match text_year > "freqEnwordYear.txt";
```

영어 슬로건의 출현빈도는 산업별로도 차이를 보이는데 다음 검색식을 통해 언어와 산업간의 상호작용에 대한 통계를 산출해낼 수 있다. 분석결과 컴퓨터 산업과 의류 산업 및 전자 산업에서 영어어휘가 많이 사용되는 것을 확인할 수 있다. 이 점은 이러한 산업들이 글로벌 시장에서 비즈니스를 하는 환경과 무관하지 않은 것으로 보인다.

다른 한편, 영어의 사용이 브랜드와도 상관성을 가질 수 있는데, 코퍼스의 분석을 통해 우리는 “Adidas”, “Coca-Cola” 등 다국적 기업에서 생산되는 브랜드가 광고에서 영어를 많이 사용한다는 사실을 다시 확인할 수 있다.

광고슬로건을 언어학적으로 연구할 때 관심을 가져볼 만한 주제 중의 하나는 슬로건의 길이와 관련한 분포이다. 이 때 길이의 단위는 토큰이 되고, 코

퍼스를 기반으로 한 토큰 개념속에는 ‘,’ ‘!’ 및 ‘?’ 등 문장부호도 함께 계산된다. 다음 [표 6]은 코퍼스 GLOS에 포함된 슬로건의 길이에 대한 분포를 보여준다.

길이	빈도
1	2
2	95
3	1,480
4	3,704
5	3,217
6	2,982
7	1,589
8	1,030
9	626
10	370
11	229
12	173
13	102
14	54
15	39
16 이상	67
합계	15,759

<표 6> 슬로건의 길이에 대한 분포

이 표를 보면 길이가 4인 슬로건 수가 가장 많고 길이가 5인 슬로건이 큰 차이 없이 뒤따르고, 슬로건 길이 6이 제3위를 차지하고 있다.

1980년대의 광고슬로건 750개를 연구한 Baumgart (1992: 69)가 제시한 다음 표에 따르면, 단어수가 4개에서 6개에 이르는 슬로건 수가 60%을 넘는다.

길이	빈도	백분율
2-3	178	23.7%
4-6	464	61.9%
7-9	94	12.5%
10-12	12	1.6%
13-14	2	0.3%
합계	750	100%

<표 7> 1980년대 광고슬로건의 길이 분포

GLOS의 길이에 대한 분포를 Baumgart의 연구결과와 비교하기 위하여 위 [표 7]의 통계를 구간별로 나누어 제시하면 다음 [표 8]과 같다. Baumgart에서와 달리 코퍼스 GLOS에서는 문장부호도 토큰 하나로 간주되기 때문에 구간을 나눌 때에, Baumgart의 경우보다 하나씩을 더했다.

길이	빈도	백분율
1-4	5,281	33.5%
5-7	7,788	49.4%
8-10	2,026	12.9%
11-13	504	3.2%
14 이상	160	1.0%
합계	15,759	100%

<표 8> GLOS에서의 슬로건 길이 분포

이 분포를 보아도 역시 5개 토큰 이상 7개 토큰 이하 구간이 가장 많은 것을 알 수 있다.¹¹⁾ 그러나 이 구간도 Baumgart의 경우 만큼 집중도가 높지는 않은 것에 주목할 필요가 있다.

이러한 통계를 얻기 위해 사용하는 검색식은 다음과 같다.

11) Janich (2001: 48)에 따르면, 슬로건의 기능은 “제품, 브랜드 및 회사의 재인지”에 있다. 슬로건들이 대부분 7개 단어를 넘지 않은 것은 기억의 용이성과 관련된다.

```
(25) T=<s>[pos=".*"]{5,7}</s>;
      count by word > "freqSentlength5_7.txt";
```

이 검색식을 이용하면 토큰이 5개 이상 7개 이하로 구성된 슬로건들을 빈도와 함께 용례로 추출할 수 있다. 아래에 결과의 일부가 제시된다.

```
(26) 빈도   용례
      4     Das Leben ist schön .
      4     Die neue Sicht der Dinge .
      4     Eine Klasse für sich .
      3     Das Plus von Siemens .
      3     Das haben Sie noch nie gesehen .
```

여기에서 확인할 수 있듯이 문장부호인 ‘.’가 토큰 하나로 계산되어 있다. 첫 열의 숫자는 해당 슬로건 자체의 출현빈도이다. 이 빈도들을 모두 합하면, 토큰 수가 5개 이상 7개 이하인 슬로건들의 수 전체를 구할 수 있다. 합산결과는 7,788개이다.

Baumgart (1992: 70)에서는 슬로건에 나타난 어휘들의 품사분포에 대해서도 논의하고 있는데, 그녀에 따르면 명사가 가장 많이 나타나는 것으로 분석되었다. 다음 데이터를 보자.

품사	빈도	백분율
명사 Substantive)	541	72.1%
동사 Verben	272	36.3%
형용사 Adjektive	372	49.6%
합성어 Komposita	89	11.9%
상표명 Markennamen	284	37.9%
합계	750	100%

<표 9> Baumgart (1992)에서의 품사분포

코퍼스 GLOS에서 추출한 이와 유사한 통계는 다음과 같다.

품사	빈도	백분율
명사 (NN)	20,143	22.24
동사 (V.*)	9,607	10.62
형용사 (ADJA/ADJD)	8,560	9.45
고유명사 (NE)	2,461	2.71
기타	49,796	54.98
합계	90,567	100%

<표 10> 코퍼스 GLOS의 품사분포

이 표에서 우리는 코퍼스 GLOS에서도 명사가 차지하는 비율이 매우 높지만 집중도에서 있어서는 Baumgart의 경우와 큰 차이를 보인다는 사실을 확인할 수 있다. 이 까닭은 코퍼스 GLOS의 경우 품사 태깅이 되어 있는 코퍼스로서 품사에는 명사, 형용사, 동사 등 내용어 뿐만 아니라 문장부호나 관사 등 문법기능을 수행하는 토큰들도 포함되기 때문에 내용어의 비중이 상대적으로 약화된 데에 기인한다.

코퍼스 GLOS를 통해 검증 가능한 이슈 중의 하나는 품사연쇄의 분포에 관한 것이다. 코퍼스로부터 추출한 분포의 일부가 아래의 표에 제시된다.

순위	품사연쇄유형	빈도	누적 빈도	누적 백분율 (%)
1	ART ADJA NN \$.	440	440	2.79
2	NN APPR NN \$.	324	764	4.85
3	ART NN \$.	268	1,032	6.55
4	ART NN ART NN \$.	138	1,170	7.42
5	ADJA NN \$.	129	1,299	8.24
6	NN APPR ART NN \$.	120	1,419	9

7	ADJD ADJD \$.	113	1,532	9.72
8	ART NN APPR NN \$.	105	1,637	10.39
9	ART NN APPR ART NN \$.	82	1,719	10.91
10	NN APPR ADJA NN \$.	81	1,800	11.42

760	ADJD \$, KOUS PPER VVFIN	3	8,044	51.04
761	2 ADJA \$(NN \$.	2	8,046	51.44

7826	WRB VBP PP VBP TO VB PP \$ NN SENT	1	15,759	100

<표 11> GLOS에서의 품사연쇄의 분포

이 분포를 살펴보면 빈도 순위 1위부터 상위 10% 정도에 해당하는 순위 760위의 품사연쇄유형들이 차지하는 누적백분율이 51.04%에 달한다. 그런데 이러한 결과는 언어학의 파레토 Pareto 법칙으로 불리는 Zipf의 언어경제성 원리와 들어맞지 않은 것으로 풀이된다.¹²⁾ 왜냐하면, 대부분의 언어학적인 분포는 그것이 어휘 층위이든, 품사 층위이든 구구조규칙 층위이든 출현빈도 상위 10%가 80%이상의 누적백분율을 기록하기 때문이다 (이민행 2012: 468). 뒤집어 생각하면, 이 사실은 광고가 독창적이고 창의적이라는 일반적인 인식을 뒷받침하는 것일 수도 있다.

3. 맺음말

지금까지 우리는 광고슬로건 코퍼스 GLOS의 활용방안에 대해 논의를 하였다. 이 코퍼스는 검색엔진 CQP를 장착하고 있어 다양한 유형의 검색이 가능하다. 논의를 통해 광고슬로건의 여러 가지 언어적 특성들을 밝혀냈다. 이 특

12) 한 심사위원의 의견대로, 품사연쇄를 넘어 NP, ADJP 등과 같은 구범주연쇄를 분석한다면 광고슬로건도 Zipf의 법칙을 따른다는 사실을 밝혀낼 수 있을 지도 모른다. 그러나, 이러한 분석작업은 전체 슬로건에 대한 구문분석 데이터를 바탕으로 해야 하기 때문에 후속연구를 기대해야 할 것이다.

성들을 열거하면 다음과 같다.

- 명사의 시대별 사용양상 (1950년대-2000년대)
- 접미사 -bar의 파생어들 목록
- 접미사 -wert의 파생어들 목록
- 부사 - 형용사 연속체 구문
- 결과구문의 사용양상
- 문장유형의 분포
- 시대별 슬로건들 간의 문서유사도
- 산업별 슬로건들 간의 문서유사도
- 영어의 시대별 사용양상
- 영어의 브랜드별 사용양상
- 문장길이의 분포
- 품사의 분포
- 품사연쇄의 분포

참고문헌

- 김원식 (2001): 광고 슬로건. 『言語와 言語學』 26, 1-34. 한국외국어대학교 언어 연구소.
- 오장근 (1999): 광고텍스트의 전략적 이해. 『독어학』 1, 287-311.
- 이민행 (2012): 『전산 구문문법 연구』. 도서출판 역락.
- 정동규 (2000): 광고언어의 단어유희. 『독어학』 2, 185-206.
- Baumgart, M. (1992): Die Sprache der Anzeigenwerbung. Eine linguistische Analyse aktueller Werbeslogans. Heidelberg: Phisica-Verlag.
- Evert, S. (2010): cqP manual. The IMS Open Corpus Workbench (CWB) CQP Query Language Tutorial. - CWB Version 3.0. University Stuttgart.
- Gries, S. Th. (2008): Statistik für Sprachwissenschaftler. Göttingen: Vandenhoeck & Ruprecht.
- Hardie, A (2012): CQPweb - combining power, flexibility and usability in a

- corpus analysis tool. In: International Journal of Corpus Linguistics 17(3), 380-409.
- Hennemann, A. /Schlaak, C. (Hrsg.) (2013): Korpuslinguistische Untersuchungen. Analysen einzelsprachlicher Phänomene. Berlin: Frank & Timme.
- Janich, N. (2001): Werbesprache. Ein Arbeitsbuch, 2. Auflage. Tübingen: Gunter Narr Verlag.
- Römer, R. (1976): Die Sprache der Anzeigenwerbung. Düsseldorf.

Zusammenfassung

Untersuchungen zu den sprachlichen Charakteristika der deutschen Werbeslogans

- Ein korpuslinguistischer Ansatz

Lee, Min-haeng (Yonsei Univ.)

In diesem Beitrag wird diskutiert, wie man ein Korpus von Werbeslogans zum Zwecke linguistischer Untersuchungen nutzbar machen kann. Das Korpus, das in der Forschung herangezogen wird, heißt GLOS (German Language Of Slogans). Es besteht aus 90.567 Tokens (15.757 Sätzen). Das Korpus ist mit der Suchmaschine CQP(corpus query processor) ausgestattet. Sie bietet also die Möglichkeit, in dem Korpus zu recherchieren.

Durch die Untersuchung lassen sich folgende Erkenntnisse erwerben:

- Verwendungstendenz der Nomina nach dem Zeitabschnitt
- Adverb-Adjektiv Wortfolge-Konstruktion
- Verwendungsweise der sogenannten Resultativ-Konstruktion
- Verwendungstendenz von englischen Wörtern
- Ähnlichkeiten von Dokumenten
- Distribution der Satzlänge
- Distribution der Wortarten
- Distribution der Wortartenfolgen

[검색어] 코퍼스, 광고슬로건, 문서유사도, 사용경향, 품사의 분포, 문장길이
Korpus, Werbeslogan, Ähnlichkeit der Dokumente, Verwendungstendenz,
Distribution der Wortarten, Satzlänge

이민행 120-749
서울시 서대문구 신촌동 연세로 50
연세대학교 문과대 독어독문학과
leemh@yonsei.ac.kr

논문 접수일: 2014. 04. 28
논문 심사일: 2014. 05. 25
게재 확정일: 2014. 06. 10