

신문 기사의 코퍼스 언어학적 분석

송경화 · 강범모
고려대학교 언어학과(고려대학교 민족문화연구원)

Corpus-Linguistical Analysis of Newspaper Articles

Kyung-hwa Song, Beom-mo Kang
Department of Linguistics, Korea University, Seoul, Korea
(Institute of Korean Culture Researcher, Korea University, Seoul, Korea)

요 약

본 연구에서는 신문 기사에 대한 실증적 언어 분석을 목적으로 하여, <21세기 세종계획>에 의해 구축된 대용량의 신문 기사 말뭉치를 다양한 각도로 계량화하여 분석한다. 신문 기사를 표제, 전문, 본문의 구성으로 나누고 각 구성의 특징에 따라 형태 분석 말뭉치, 형태의 미 분석 말뭉치, 구문 분석 말뭉치를 이용하여 분석한다. 본 연구는 대량의 신문 기사 말뭉치를 이용한 계량적 방법이라는데 의의가 있다. 이러한 연구 방법을 통하여 기존의 직관을 이용한 연구 방법들과 차별화된 실증적 연구로서 신문 이론을 검증하고, 신문 기사의 새로운 언어 현상을 발견할 수 있을 것이다.

1. 서론

기존의 신문 기사에 대한 연구들은 경험적인 언어 분석이 많이 이루어지지 못했던 것이 사실이다. 신문학 분야의 연구는 기사 스타일 및 작성법에 관한 연구들이 대부분이며, 실증적인 연구로는 특정 기사의 내용에 대한 비교분석이 주를 이룬다. 1) 언어학 분야의 신문 기사 연구는 실증적인 연구이기는 하지만, 텍스트의 언어 특성을 연구하면서 일부 장르로서만 다루어졌기 때문에 2) 신문 기사에 대한 깊이 있는 분석적인 연구라고는 볼 수 없다.

국어학 분야에서의 신문 기사 연구는 주로 담화 형식에 의한 문체론의 한 분야로 다루어지고 있다. 박갑수(1998)에서 신문의 문체와 표현에 대한 실증적인 분석이 시도되었다. 이 연구는 이전까지 이론적인 연구가 대부분이었던 것을 감안하였을 때 새로

운 시도로서 큰 의의를 갖는다. 그러나 분석 자료가 특정 기간에 몰려있고, 적은 분량의 텍스트를 이용하였기 때문에 일반적인 특성을 파악하기에 부족함이 없지 않다. 또한, 자료에 대한 정확한 분량의 제시와 분석 방법에 대한 구체적인 정보가 부족하다고 할 수 있다.

본 연구에서는 기존 연구들과 차별된 방식으로, 제시된 이론들의 검증과 함께 신문 기사의 새로운 특성을 파악하는 것을 목표로 한다. <21세기 세종계획>에 의해 구축된 대용량의 신문 기사 말뭉치를 다양한 각도로 계량화하여 검토함으로써 기존의 신문 이론을 검증하거나, 새로운 현상을 발견할 수 있을 것이다.

2. 연구 대상

신문 기사는 보통 표제(제목: headline), 전문(서두: lead), 본문(body)의 세 부분으로 구성된다.

표제는 신문 기사의 머리에 있는 것으로, 독자들에게 이어지는 기사의 주요한 정보를

1) 박재영·신명선(2004), 박종화(2003)

2) 강범모·김홍규·허명희(2000)

주려고 요약 표현한 것이다.

전문은 전달하려는 전체 내용을 요약한 것이며, 역피라미드형과 혼합형의 신문 기사에서 첫 번째 문단으로 쓰인다. 신문 기사의 대부분이 역피라미드형인 점을 감안하였을 때, 해설 기사나 사설을 제외한 거의 모든 기사에 전문이 포함되어있다고 볼 수 있다.

본문은 전문을 부연하고 고증하는 부분이다. 즉, 표제와 전문에서 드러난 기사의 요지 또는 기사 주제에 대한 주요한 사실들을 보충하며, 사건과 관련한 흥미로운 사실을 기술한다.

신문 기사는 그 내용과 목적에 따라 크게 보도-해설 기사와 사설-칼럼 기사로 나눌 수 있다. 사건을 객관적으로 기술하는 보도-해설 기사와 필자의 개인적인 의견을 담은 사설-칼럼 기사는 그 내용과 형식면에서도 차이를 보인다.

본 연구에서는 신문 기사를 표제 전문 본문의 구성으로 나누어 살펴보고, 필요에 따라 보도-해설 기사와 사설-칼럼 기사의 구분을 두고 각각의 구성에 대한 적절한 방법의 계량적 연구를 이용할 것이다.

본 연구는 <21세기 세종계획>에서 구축한 국어 기초자료 말뭉치를 분석 대상으로 삼는다. 1998년부터 2004년까지 7년에 걸쳐 구축된 말뭉치는 원시 말뭉치, 형태분석 말뭉치, 형태미분 분석 말뭉치, 구문분석 말뭉치로 나뉜다.

본 연구의 표제과 전문 분석에서는 형태 분석 말뭉치를, 본문 분석에서는 형태미분 분석 말뭉치와 구문분석 말뭉치를 이용할 것이며, 각 말뭉치의 신문 기사 자료를 분석 대상으로 할 것이다.

각각의 말뭉치 규모와 신문 기사 비율을 표로 정리하면 다음과 같다.

구분	구축 어절 수	신문 기사 어절 수	신문 기사 비율
원시 말뭉치	62,429,781	17,884,995	28.6%
형태분석 말뭉치	12,131,018	2,676,699	22.0%
형태미분 분석 말뭉치	8,109,436	1,713,995	21.1%
구문분석 말뭉치	363,226	93,798	25.8%

표 6 <21세기 세종계획> 국어 기초자료 말뭉치

신문 기사의 구성별 특성 혹은 장르적 특성을 분석하기 위해서는 비교 말뭉치가 필

요하다. 비교 말뭉치는 <21세기 세종계획> 국어 기초자료 말뭉치에서 신문을 제외한 일반 말뭉치들 중 적절한 장르 비율을 고려하여 구성하였다.

3. 표제 분석

신문 기사 말뭉치 2,506,665어절에서 추출한 표제의 총 어절 수는 115,492어절이다. 일반 말뭉치와의 비교를 위해 같은 어절 수의 비교 말뭉치를 선정하여 비교분석하였다.

3.1 표제의 특성

신문의 표제는 1)표시성, 2)압축성, 3)감동성, 4)품위, 5)심미성의 기능을 지닌다.³⁾ 본고에서는 표제의 ‘표시성’과 ‘압축성’이라는 특성에 주목하여, 이러한 특성들이 실제 말뭉치에서 어떠한 방식으로 표현되었는지를 분석한다.

표시성

신문의 표제는 내용을 드러낸다. 문학 장르의 경우 그 제목이 상징적인 경우가 많지만, 신문 표제의 경우는 표제만 보고도 그 내용을 알 수가 있다. 그렇기 때문에 신문 표제에 있어서 표시성이라는 것은 상당히 중요한 기능을 한다. 짧은 표현으로 독자들에게 명료한 정보를 주기 위해 표제는 다음과 같은 특성들을 나타낸다. 다음의 분석은 신문 기사 표제와 비교 말뭉치의 형태별 토큰 개수를 비교한 결과이다.

	표제		일반	
	형태 수	비율	형태 수	비율
일반-고유명사	107,220	52%	59,67	23%
대명사	441	0.2%	4,002	1.5%
종결부호	596	0.29%	8,761	3.39%
, · : ; /	4,947	2.40%	4,478	1.73%
인용부호	18,145	8.82%	5,060	1.96%
~	286	0.14%	79	0.03%
기타기호	2,563	1.25%	144	0.06%

표 7 표제/일반 말뭉치 비교 1

신문 기사 표제의 경우, 일반 말뭉치보다 명사의 쓰임이 훨씬 많은 것을 볼 수 있다. 이것은 표제가 수식어 등의 군더더기 없이, 명사를 사용하여 내용을 요약함으로써, 최

3) 하타노칸지(1966: p.226-234)

소한의 표현으로 많은 정보의 정확한 전달을 목적으로 함을 보여준다.

명사와는 반대로 대명사의 경우 표제에서의 쓰임이 적다는 것을 알 수 있다. 이것은 지시 대상의 모호성을 피하기 위함이라고 생각할 수 있다.

기호의 경우, 종결 부호를 제외한 나머지의 기호들이 표제에 더 많이 쓰이고 있음을 알 수 있다. 언어 대신 기호를 사용함으로써 이해 속도를 높이고, 특히, 인용부호를 많이 사용하여 독자의 흥미와 관심을 유발할 수 있을 것이다.

압축성

신문의 표제는 그 자체가 하나의 기사로서 사건 자체의 압축된 기술이어야 한다. 바쁜 독자와 한정된 지면을 고려한 압축이 기사의 주요한 특징이 되는 것이다. 압축의 대표적인 방법은 생략이다. 다음의 분석은 신문 기사 표제 말뭉치와 비교 말뭉치의 형태별 토큰 개수를 비교한 결과이다.

	표제		일반	
	형태 수	비율	형태 수	비율
조사	13,160	6%	45,127	17%
수식언	3,867	2%	12,551	5%
말줄임표	1,674	0.81%	174	0.07%

표 8 표제/일반 말뭉치 비교 2

신문 표제 말뭉치는 일반 말뭉치와 비교해 조사의 쓰임이 훨씬 적은 것을 볼 수 있다. 조사를 생략함으로써 빠르고 간결하게 정보를 전달할 수 있을 것이다.

수식언(관형사, 부사) 또한 표제에서 적게 나타남을 알 수 있다. 이것은 앞서 살펴본 표시성의 분석에서, 표제의 명사 출현빈도가 높다는 것과 관계가 있다. 수식언의 출현빈도가 낮아짐으로써 명사의 출현빈도가 높아지는 것이다. 최소한의 표현으로 최대한의 정보를 전달하기 위해 부가어가 생략되는 경우가 많다.

표시성 분석에서 살펴본 기호의 높은 출현빈도와 관계하여 신문 기사 표제에 말줄임표가 매우 높은 빈도로 출현함을 알 수 있다. 말줄임표를 이용해 필요 없는 내용을 생략할 수 있다.

다음은 표제의 종결형을 분석한 것이다.

	어절 수	비율
완전 종결	1,504	6%
불완전 종결	24,861	94%
총	26,365	100%

표 9 표제 종결형

박갑수(1998: p.14)에서 표제는 기사의 요약으로 정확하고 구체적인 낱말을 써 이해하기 쉽게 쓴 완전한 진술 문장이라고 하였다. 이론적으로 표제는 주술어를 갖춘 완전한 문장으로서 진술의 임무를 수행하게 되는 것이다. 그러나 위의 분석에서 볼 수 있듯이, 우리나라 신문 기사의 표제는 완전한 진술 문장보다는 불완전한 진술 문장이 압도적으로 많다.

이러한 특성은 표제와 일반 말뭉치의 용언에 대한 빈도 분석 결과에도 나타나 있다.

	표제		일반	
	형태 수	비율	형태 수	비율
용언	10,555	5%	36,960	14%

표 10 표제/일반 말뭉치 용언 비교

이러한 결과 역시 압축성이라는 신문 표제의 특성에 기인한 것이라고 볼 수 있다. 불완전 종결의 경우, 종결 어미 대신 어떤 성분들로 문장의 끝을 맺게 되는지 살펴보면 다음과 같다.

	어절 수	비율
체언	19,990	80.4%
연결어미	2,453	9.9%
조사	1,315	5.3%
어근	627	2.5%
부사	371	1.5%
전성어미	105	0.4%
총	24,861	100%

표 11 표제 불완전 종결 구성

표제는 내용과 형식이 완전한 초행완전(初行完全)을 원칙으로 하고 있으나, 4) 실제로 94%의 표제가 불완전한 형식임을 확인할 수 있다. 또한, 기사 작성 요령을 설명하는 많은 문헌들에도 이미 불완전 종결 표제를 당연시하고 있다. 5) 이러한 사실들로 미루어 한국 신문 기사에서는 표제 초행완전이라는 원칙이 지켜지고 있지 않음을 확인할 수 있다.

4) 한국언집기자회(1978: p.196)

5) 박종화(2003)

3.2 보도-해설 표제와 사설-칼럼 표제의 비교

보도-해설과 사설-칼럼은 그 내용과 형식에 있어서 세부 장르로도 나뉠 수 있을 만큼의 차이가 있으므로, 어떻게 다른지 분석해볼 필요가 있다.⁶⁾

	보도-해설		사설-칼럼	
	형태 수	비율	형태 수	비율
채언	102,935	60%	12,290	45.4%
용언	8,217	4.8%	2,292	8.4%
수식언	3,107	1.8%	723	2.6%
조사	9,737	5.6%	3,130	11.6%
어미	10,181	5.9%	3,241	12%
어근	1,408	0.8%	239	0.8%
기호	22,540	13.1%	3,315	12.2%
기타	13,445	8%	1,797	7%
총 형태 수	158,125	100%	25,230	100%

표 12 보도-해설/사설-칼럼 표제 비교

위의 표와 같이 보도-해설과 사설-칼럼의 표제 비교 결과는 신문 기사 표제 전체와 일반 말뭉치의 비교 결과와 유사하다. ‘표시성’과 ‘압축성’이라는 것은 정보의 빠른 전달과 명확한 이해를 돕기 위한 기사 표제의 특성인데, 사설과 칼럼이라는 것은 빠른 정보 전달과는 약간의 거리가 있기 때문에 보도-해설 표제에서 보여진 ‘표시성’과 ‘압축성’의 요인들이 적게 나타나는 것이다.

이러한 특성은 보도-해설 표제의 종결형과 사설-칼럼 표제의 종결형을 비교해보아도 알 수 있다.

	보도-해설	사설-칼럼
완전 종결	846 (3.8%)	849 (21.8%)
불완전 종결	21,625 (96.2%)	3,045 (78.2%)
총	22,471 (100%)	3,894 (100%)

표 13 보도-해설/사설-칼럼 표제 종결형 비교

위의 표에서 보듯, 보도-해설 표제보다 사설-칼럼 표제의 완전 종결 비율이 더 높다. 이것은 사설-칼럼의 표제가 불완전 종결을 이용한 압축의 기능이 적다는 것을 말

6) 형태분석 말뭉치 신문 자료는 보도-해설 기사가 1,605,901어절(64%), 사설-칼럼 기사가 900,764어절(36%)을 차지한다. 그 중 표제의 비율은 보도-해설 표제 100,751어절(87%), 사설-칼럼 표제 14,741어절(13%)이다.

해준다. 또한 사설-칼럼의 내용적 특성상 독자에게 전달하고자 하는 바를 종결형에 담아내는 경우가 많기 때문일 것으로 해석해 볼 수도 있다.

4. 전문 분석

전문 분석에서는 전문과 일반 말뭉치의 형태별 토큰 개수를 비교하여 전문이라는 구성이 다른 것들과 어떠한 차이가 있는지 살펴본다.

신문 기사 말뭉치 2,506,665어절에서 추출한 전문의 총 어절 수는 266,672어절(11%)이다. 일반 말뭉치와의 비교를 위해 같은 어절 수의 비교 말뭉치를 선정하였다.

4.1 전문의 특성

이 절에서는 신문 기사의 전문과 비교 말뭉치의 형태별 토큰 개수를 비교하여 전문이 가진 특성을 살펴보고자 한다.

먼저 일반-고유명사 형태 수의 비교 결과는 다음과 같다.

	전문		일반	
	형태 수	비율	형태 수	비율
일반-고유명사	196,234	34%	139,750	23%
대명사	2,329	0.4%	9,259	1.5%
용언	51,293	9%	84,193	14%
수식언	12,708	2%	28,736	5%
조사	93,558	16%	105,377	18%
어미	82,538	14%	118,420	20%
기호	68,428	12%	49,640	8%

표 14 전문/일반 말뭉치 비교

전문의 경우, 일반 말뭉치보다 명사의 쓰임이 많은 것을 볼 수 있다. 이것은 3.1절에서 확인하였던 표제의 특성과 비슷한 결과이다. 전문의 명사 쓰임은 표제보다 많지는 않지만, 표제에서 나타나는 표시성의 특성을 어느 정도 갖고 있음을 보여준다. 대명사 역시 표제와 일반 말뭉치의 차이와 비슷한데, 글의 맨 앞에 위치한 전문의 특성으로 지시 대상을 명확하게 표현하기 위한 결과라고 볼 수 있다. 용언, 수식언, 조사, 어미는 일반 말뭉치에 비해 적게 쓰였고, 기호는 일반 말뭉치에 비해 많이 쓰인 것을 확인 할 수 있다.

이 결과는 정도의 차이를 제외하고, 표제와 일반 말뭉치와의 차이와 일치하는 결과이다. 즉, 신문 기사의 전문은 표제가 가지

고 있는 ‘표시성’ 과 ‘압축성’ 의 특성을 어느 정도 가지고 있다고 볼 수 있다.

그러나 전문과 일반 말뭉치의 문장 길이를 비교함으로써 앞의 분석과 매우 다른 결과를 볼 수 있다. 다음은 전문과 본문, 일반 말뭉치의 평균 문장 길이를 비교한 결과이다.

	한 문장당 평균 어절 수
전문	17.0
본문	15.7
일반 말뭉치	13.0

표 15 전문/본문/일반 말뭉치 평균 문장 길이 비교

위의 표는 전문이 표제와 비슷하게 ‘압축성’ 의 특성을 갖는다는 측면에서 의외의 결과를 보여준다. 문장 내에서 나타난 ‘압축’ 이라는 특성이 문장 길이로 표현되지는 않은 것이다. 본문 역시 일반 말뭉치에 비해 문장의 길이가 길다는 것을 알 수 있다.

전문은 독자로 하여금 그 문장에 시선을 멈추고 기사를 읽도록 관심을 유발해야 한다. 그러기 위해서는 사건의 핵심을 꼬집어 낸 간결하고 날카로운 진술이 필요하다. 그러나 위의 결과는 전문이 이러한 기능을 제대로 하지 못하고 있음을 나타낸다. 수식언과 조사 등이 적게 사용되고, 명사가 많이 사용됨으로써 문장 내 압축의 기능이 높은 반면, 가독성(readability)의 기능은 떨어진다고 볼 수 있다. 이것은 전문을 쓸 때 육하원칙의 표현과 독자의 흥미유발 및 사건에 대한 빠른 인식을 위하여 하나의 문장에 많은 내용을 담으려 했기 때문인 것으로 해석해볼 수 있다.

4.2 보도 기사의 표제와 전문

이전 분석에서도 보았듯이 표제와 전문은 비슷한 기능과 특성을 갖고 있다. 그렇다면, 같은 기사 안에서의 표제와 전문은 그 내용을 이끌어 얼마나 유사성을 보이는지 살펴볼 필요가 있을 것이다. 여기에서는 표제와 해당 전문을 하나의 세트로 설정해 내용을 중심으로 동일 형태 비율을 알아볼 것이다. 자동화의 문제로 인해, 약 8,000개의 기사 중 100개의 기사를 선별하여 분석하였다. 선별된 기사는 다음과 같다.

신문사	기사 수	장르
동아	20	(외신, 사회)
조선	20	(경제, 사회)
한겨레	20	(경제, 사회)
중앙	10	(과학)
한국	10	(경제)

표 16 표제-전문 세트 구성

먼저, 내용어를 대상으로 한 표제와 전문의 기사당 평균 형태 수는 다음과 같다.

	표제	전문
일반명사	10.19	19
고유명사	1.16	2.82
동사	0.65	2.93
형용사	0.14	0.49
총	12.14	25.24

표 17 표제/전문 기사당 평균 형태 수 비교

표제와 전문은 서로 비슷한 기능을 갖기 때문에 일반적으로 표제의 어휘가 전문에 그대로 사용되는 경우가 많을 것이라고 예상할 수 있다. 이러한 추측을 확인하기 위해 표제와 전문에서 동시에 사용된 동일 형태를 분석해 보았다. 그 결과, 표제를 기준으로 했을 때, 표제 안의 36.5%의 내용어가 전문에 사용되었으며, 전문을 기준으로 했을 때, 전문 안의 17.7%의 내용어가 표제의 것을 그대로 따랐다. 즉, 표제와 전문의 동일 형태 비율이 높은 것으로 분석되었다. 동일 어휘 이외에 유사어나 약어 등의 사용을 함께 고려하여 분석한다면 더 높은 결과가 나올 수도 있을 것이다. 본 연구에서는 유사어나 약어까지 동일 형태 범주에 포함시키지 않았으나, 실제로 많은 형태의 유사어나 약어가 사용되었다.

4.3 사실-칼럼 표제와 전문

3.2절에서 보도-해설의 표제보다 사실-칼럼의 표제가 일반 말뭉치와 비슷한 특성을 가지고 있음을 확인하였다. 4.1절에서는 전문을 일반 말뭉치와 비교하여 그 특성이 표제와 비슷함을 살펴보았다. 그렇다면, 표제의 특성을 가진 전문과 일반 말뭉치의 특성을 가지고 있는 사실-칼럼 표제가 서로 비슷한 특성을 갖고 있을 것이라는 추측이 가능하다.

사실-칼럼의 표제와 전문이 유사한 특성을 갖거나 혹은 그렇지 않다는 것을 증명하기 위해 보도-해설 표제와 일반 말뭉치를

함께 비교할 필요가 있다. 보도-해설 표제와 사실-칼럼 표제는 3.2절의 결과를 사용하였으며, 전문과 비교용 일반 말뭉치는 보도-해설 표제의 어절 수에 맞추어 구성하였다.

다음은 보도-해설 표제와 사실-칼럼 표제, 전문, 일반 말뭉치를 함께 비교한 결과이다.

	보도	사실	전문	일반 말뭉치
	형태 수 (비율)	형태 수 (비율)	형태 수 (비율)	형태 수 (비율)
체언	102,935 (60.0%)	12,290 (45.4%)	89,143 (42.3%)	63,002 (27.9%)
용언	8,217 (4.8%)	2,292 (8.4%)	16,685 (8.0%)	32,355 (14.3%)
수식언	3,107 (1.8%)	723 (2.6%)	3,806 (1.8%)	11,054 (4.9%)
조사	9,737 (5.6%)	3,130 (11.6%)	33,696 (15.0%)	39,178 (17.3%)
어미	10,181 (5.9%)	3,241 (12.0%)	28,848 (13.7%)	45,175 (20.0%)
어근	1,408 (13.1%)	239 (12.2%)	893 (7.7%)	2,167 (7.7%)
기호	22,540 (0.8%)	3,315 (0.8%)	16,222 (0.5%)	17,349 (0.9%)
기타	13,445 (8.0%)	1,797 (7.0%)	21,701 (10.0%)	15,054 (7.0%)
총	171,570 (100%)	27,027 (100%)	210,994 (100%)	225,334 (100%)

표 18 장르별 비교

위의 결과에서 보듯이, 수식언을 제외한 모든 부분에서 사실-칼럼의 표제와 전문이 비슷한 특성을 가지고 있음을 확인할 수 있다. 사실-칼럼의 경우, 수식언의 쓰임이 많은 것은 사실-칼럼의 표제가 압축의 특성이 덜하고, 화려한 수식어의 사용으로 독사의 관심을 유발시키고자 하는 의도로 해석해볼 수 있다.

5. 본문 분석

표제와 전문을 제외한 본문은 기사의 알맹이로서 신문의 일반적인 장르 특성을 가장 잘 보여준다고 할 수 있다. 그러므로 이 장을 통하여 신문 기사의 일반적인 특성을 파악할 수 있다.

5.1 문장 구조 분석

이 절에서는 93,798어절 규모의 신문 기사 구문분석 말뭉치를 이용하여 단문과 복문의 쓰임을 분석한다. 일반 말뭉치와의 비교를 위해 같은 규모의 비교 말뭉치를 선정

하였다.

	본문		일반 말뭉치	
	형태 수	비율	형태 수	비율
단문	3,477	56%	5,902	67%
복문	2,783	44%	2,944	33%
총 문장 수	6,260	100%	8,846	100%

표 19 본문/일반 말뭉치 문장 비교

본문과 일반 말뭉치의 단문과 복문 개수는 위와 같다. 신문 기사의 본문은 구조가 단순한 문장으로 구성되어야 한다. 그러나 위의 결과에서 보듯이 신문 기사의 본문이 일반 말뭉치에 비해 구조가 더욱 복잡하다는 것을 알 수 있다.

본문과 일반 말뭉치의 복문 구성은 다음과 같다.

	본문		일반 말뭉치		
	형태 수	비율	형태 수	비율	
접속	1,056	29%	1,172	32%	
내포	명사절	159	4%	169	4%
	관형사절	1,889	52%	1,801	49%
	부사절	83	2%	76	2%
인용절	462	13%	480	13%	
총	3,649	100%	3,698	100%	

표 20 본문/일반 말뭉치 복문 구성 비교

접속과 내포의 관형사절을 제외하고 비슷한 비율로 복문이 구성됨을 알 수 있다. 신문 기사 본문은 일반 말뭉치보다 접속의 구성이 적으며, 반대로 내포의 관형사절 구성이 많다.

복문 구성을 살펴봄으로써, 신문 기사 전체와 일반 말뭉치의 많은 차이를 발견할 수는 없다. 그러나 일반 문어의 복문에 있어서 접속보다는 내포의 구성이 많이 쓰이며, 내포의 구성 중 관형사절이 다른 절에 비해 월등히 많이 쓰이고 있다는 것을 분석을 통하여 확인해 볼 수 있었다.

5.2 어휘 분석

이 절에서는 1,713,995어절의 신문 기사 형태미분석 신문 기사 말뭉치를 이용하여 일반명사 중 고유어, 외래어, 한자어 등의 사용 비율을 분석해볼 것이다. 일반 말뭉치와의 비교를 위해 같은 규모의 비교 말뭉치를 선정하였다.

	본문		일반 말뭉치	
	형태 수	비율	형태 수	비율
한자어	685,767	82.8%	578,647	67.7%
고유어	85,533	10.3%	215,904	25.3%
미등록어	27,562	3.3%	20,430	2.4%
외래어	20,387	2.5%	26,430	3.1%
고유한자	7,264	0.9%	11,800	1.4%
외래한자	1,764	0.2%	1,182	0.1%
외래고유	53	0%	46	0%
고유한자의외래	1	0%	2	0%
총	828,330	100%	854,439	100%

표 21 본문/일반 말뭉치 일반명사 구성 비교

본문과 일반 말뭉치의 일반명사 구성은 위와 같다. 일반 말뭉치에 비해 신문 기사의 본문에서 한자어가 많이 쓰이며, 상대적으로 고유어의 쓰임은 적게 나타나고 있음을 볼 수 있다.

한자어와 고유어가 대부분의 일반명사를 차지하고 있기 때문에 다른 종류의 어휘들의 비율이 높지는 않지만, 외래어의 경우 신문 기사의 본문이 일반 말뭉치보다 적게 나타나는 현상은 주목할 만하다. 신문에 사용된 외래어가 사전에 등록되지 않은 경우가 많아 미등록어로 분류되었을 가능성이 있다.

표제, 전문, 본문의 일반명사 구성 분석 결과 뚜렷한 구성의 차이를 보이지 않았다. 그렇기 때문에 다음의 신문 기사 전체와 일반 말뭉치를 비교한 결과도 본문과 일반 말뭉치를 비교한 결과와 크게 다르지 않다.

	신문 기사		일반 말뭉치	
	형태 수	비율	형태 수	비율
한자어	854,349	83.0%	578,647	67.7%
고유어	102,167	9.9%	215,904	25.3%
미등록어	36,070	3.5%	20,430	2.4%
외래어	26,054	2.5%	26,430	3.1%
고유한자	8,747	0.9%	11,800	1.4%
외래한자	2,280	0.2%	1,182	0.1%
외래고유	78	0%	46	0%
고유한자의외래	2	0%	2	0%
총	1,029,745	100%	854,439	100%

표 22 신문 기사/일반 말뭉치 일반명사 구성 비교

이러한 분석은 강범모(1999)에서도 이루어진 바 있다. 한국어의 텍스트 장르와 언어 특성을 연구하면서 82가지의 언어 특성들 중 한자어와 외래어를 포함시켜 어휘의 종류를 살펴보았다. 신문 기사의 경우 2음절 이상의 한자어의 쓰임이 약 47%, 외래어의 쓰임이 약 2%로 분석되었다. 그러나, 강범모(1999)에서는 일반명사만이 아닌 모든

품사를 고려하였으며, 약 37만 어절 규모의 텍스트를 대상으로 했다는 점에서 본 연구와 차이가 있다고 할 수 있다.

신문 기사와 일반 말뭉치의 한자어, 고유어, 외래어 사용 비율을 분석한 결과는 한자어와 고유어 이외의 특이할만한 차이가 있었던 것은 아니다. 그러나 이번 분석을 통하여, 일반적인 문어에서 사용되는 일반명사의 한자어, 고유어, 외래어, 미등록어 등의 비율을 살펴볼 수 있었다는데 의의를 둘 수 있겠다.

다음은 신문 기사를 포함한 일반 문어 말뭉치의 일반명사 비율 평균을 나타낸 표이다.

	문어 말뭉치 일반 명사 비율
한자어	75.3%
고유어	17.6%
미등록어	3.0%
외래어	2.8%
고유한자	1.1%
외래한자	0.2%

표 23 문어 말뭉치 일반 명사 비율

미등록어의 경우, 여러 가지의 존재 가능성이 있다. 신조어나, 약어, 외래어, 합성어 등과 비표준어나 단순한 오타의 가능성도 있을 것이다. 본 연구에서는 미등록어에 대한 자세한 분석은 하지 않았다.

6. 결론

지금까지 신문 기사를 표제, 전문, 본문의 구성으로 나누어 각각을 다양한 각도로 분석해보았다.

신문 기사의 코퍼스 언어학적 분석은 대용량의 신문 기사 말뭉치를 이용한 계량적 방법이라는데 의의가 있다. 이러한 연구 방법을 통하여 기존의 직관을 이용한 연구 방법들과 차별된 실증적 연구로서 신문 이론을 검증하거나 새로운 현상을 발견할 수 있었다.

3.1절에서는 표제가 '표시성' 과 '압축성' 의 특성을 가지고 있다는 기존의 이론을 통해 이러한 특성들이 어떤 형태로 표출되는지 실증적으로 확인하였으며, 3.2절에서는 보도-해설 표제와 사설-칼럼 표제를 비교하여 두 표제가 형태적으로 차이가 있음을 발견할 수 있었다. 4장에서는 전문의 일반적인 특성을 확인하고, 보도-해설 표제와 해당 전문의 연관성을 파악하였으며, 사

설-칼럼 표제와 전문이 유사한 형태적 특성을 공유하고 있음을 살펴볼 수 있었다. 5.1절에서는 신문 기사가 일반 말뭉치에 비하여 문장 구조가 복잡하다는 것을 확인하였고, 복문의 세부 구성을 분석함으로써 일반적으로 복문에 있어서 접속보다는 내포의 구성이 많이 쓰이며, 내포의 구성 중에서도 관형사절이 많다는 사실을 발견할 수 있었다. 5.2절에서는 신문과 일반 말뭉치의 일반명사를 살펴봄으로써 궁극적으로, 문어에서 사용되는 일반명사의 한자어, 고유어, 외래어, 미등록어 등의 비율을 살펴볼 수 있었다는데 의의가 있다.

본 연구에서는 형태분석 말뭉치를 이용한 분석은 주로 형태별 토큰 개수를 비교하는 방법을 사용하였다. 그러나 단순히 형태별 개수를 비교하기보다는 개별적인 언어 특성에 의한 의미적 구별을 통해 더욱 구체적인 분석을 할 수 있을 것이다.

아울러 어휘 분석에서 고유어, 한자어, 외래어 이외의 사전 미등록어를 검토함으로써 어떠한 종류의 어휘가 사전에 등재되어 있지 않은지를 살펴보는 것이 하나의 과제로 남을 수 있다.

참고문헌

- [1] 강범모(1999), 한국어의 텍스트 장르와 언어특성, 서울: 고려대학교 출판부.
- [2] 강범모·김홍규·허명희(2000), 한국어의 텍스트 장르, 문체, 유형: 컴퓨터와 통계적 기법의 이용, 서울: 태학사.
- [3] 김홍규 외(1998~2004), 21세기 세종계획 국어기초자료 구축분과 연구보고서, 문화관광부.
- [4] 박갑수(1998), 신문·광고의 문체와 표현, 서울: 집문당.
- [5] 박재영·신명선(2004), 신문기사의 텍스트 언어학적 분석, 서울: 미디어 연구소.
- [6] 박종화(2003), 미디어 문장과 취재 방법론, 서울: 한울아카데미.
- [7] 한국편집기자회(1978), 신문편집, 서울: 집문당.
- [8] 하타노칸지(波多野完治)(1966), 현대문장심리학, 대일본도서.