

# 텍스트마이닝을 활용한 'ingestion\_프레임'의 한·일 'ingestibles'에 관한 일고찰

이준서\*

〈 Abstract 〉

## A Research on the Korean & Japanese 'ingestibles of 'ingestion\_frame' by Using the Text-mining Method

This paper compares the Korean and Japanese ingestibles of 'ingestion\_frame'(framenet2.icsi.berkeley.edu/) in terms of the Text-mining technical method. LEE & Han (2016) has attempted to develop the Japanese & Korean Cultural Element Mining System (CEMS). CEMS is a tool used to uncover distinctive cultural elements by comparing languages and analyzing the frequency of word co-occurrence. By way of text-mining the CEMS, we found several results.

1. 'Rice' took the first place on the frequency ranking for both Korean and Japanese.
2. On the Korean frequency ranking, 'soup'-related foods took the high places, whereas 'raw fish'-related foods took the high places on the Japanese frequency ranking.
3. Nowadays, Japanese have become more fond of 'meat'-related foods than Koreans owing to 2<sup>nd</sup> generation Koreans and the recent Korean Wave.

Field : Semantics

Keywords : Text-mining, Data-mining, Framenet, Cultural element, Ingestion\_frame

## 1. 머리말

'벼농사이론(rice theory-Talhmel Thomas, 2014-)'의 가설에 의하면 동(東)아시아인들이 관계 지향적이고 통합적인 사고를 하는 이유는 벼농사를 지었기 때문이고, 서구인들이 개인주의적이고 분석적인 사고를 하는 것은 밀농사를 지었기 때문이라고 한다. 일상생활에서 중요한 의미를 가지는 식문화는 이러한 문화적 공통점이 확연히 나타날 것으로 예측되는데, 특히 지리적·역사적으로 인접하며 언어적으로도 유사한 동일 한자 문화권에 속하여 쌀을 주식으로 삼아온 한국과 일본의 식문화에서 유사한 상호 공통분모가 발견될 가능성이 농후할 것으로 충분히 예측할 수 있다.

그러나, 이준서(2016)에서 구현된 한·일 'ingestion 프레임'의 구성요소중의 하나인 도구(instrument)에 있

\* 성결대학교 동아시아물류학부 교수

이 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-과제번호)(NRF-2019S1A5A2A03046676)

어서 한국어가 ‘숟가락’, 일본어가 ‘젓가락’이라고 하는 서로 특징적이고 대비적인 결과를 보여주고 있다. 음식을 섭취하는데 필요한 도구가 다르다는 것은 그 내용물(ingestibles)에 있어서도 차이가 있다는 것을 의미하는데, 문화적·지리적·역사적으로 상호 강한 유사성을 보이는 두 나라에서 일상생활에서 가장 기본적인 ingestion 프레임의 핵심 프레임 요소(core frame-element)에서 발견되는 대비적인 결과는 양국의 특징적인 문화적 차이라고도 할 수 있다.

본고에서는 데이터마이닝(data-mining)의 일종인 텍스트마이닝(text-mining<sup>1)</sup>) 빅데이터 분석을 통하여 한국과 일본의 각 문화권에서 섭취되는 음식물의 특성, 유행, 상호문화의 유입양상 및 영향관계를 고찰하고자 한다.

## 2. 프레임넷과 텍스트마이닝

컴퓨터가 분간할 수 있는 방대한 텍스트 기반 코퍼스(corpus)의 구축으로 1990년대 이후 자연어처리(Natural Language Processing)에 대한 관심이 증가하여 컴퓨터에 의한 자연어(natural language)의 처리가 용이해졌다. 특히 언어학의 의미론에 기반하여 구축되고 있는 어휘데이터베이스(Lexical Database)인 프레임넷(FrameNet), 워드넷(WordNet)은 전산언어학(Computational Linguistics)의 주요 관심사이기도 하다.

프레임넷의 이론적 근간이 되고 있는 틀의미론(frame semantics)이 인간의 언어와 경험, 그리고 인간을 둘러싼 백과사전적 지식(encyclopedic knowledge)을 중시하는 경험적 의미론(empirical semantics)을 바탕으로 하고 있지만, 각 언어의 다양한 문화적 변수는 그다지 고려의 대상이 되지 않았다.

틀의미론에 기반을 두면서도 다중언어의 미묘한 문화적 차이점을 찾아내기 위하여 만들어진 한·일문화소추출시스템(이준서, 2016)을 통하여 구축된 한·일문화이미지데이터베이스는 언어보편성에 치중된 틀의미론의 한계를 보완하는 것이라고 할 수 있다.

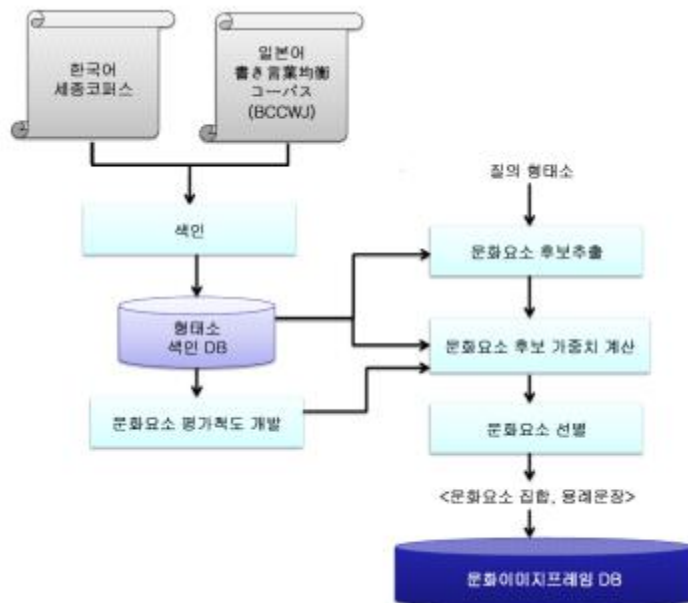
언어 속에서 문화적 요소를 찾아내기 위해서는 방대한 거대 코퍼스가 필요하며 통계적으로도 유의미한 데이터를 추출하기 위하여 데이터마이닝의 일종인 텍스트마이닝에 기반한 정확한 통계를 추출하기 위해 형태소의 품사에 따른 구분이 필요하다. 본 시스템에서는 다중언어 형태소 분석 코퍼스를 활용하였는데, 한국어 코퍼스로는 문화관광부의 ‘21세기 세종 계획 사업’의 결과물로 배포되는 세종 코퍼스를, 일본어 코퍼스로는 일본 국립국어연구소의 ‘書き言葉均衡코퍼스(BCCWJ)’를 사용하였다.

1) text mining is defined as “the process of finding of fining useful or interesting patterns, models, directions, trends, or rules from unstructured text”. Several techniques have been proposed for text mining including conceptual structure, association rule mining, episode rule mining, decision trees, and rule induction methods so far.(Un Yong Nahm & Raymond J., 2002)



〈그림 1〉 한·일 코퍼스

한·일문화요소추출시스템은 대용량 코퍼스를 효율적으로 처리하기 위해 변환하는 색인 단계, 문화요소로서 적절한지를 구분하는 평가 단계, 색인 결과와 평가척도를 적용하여 문화요소를 선별하는 검색 단계로 구성된다.



〈그림 2〉 텍스트마이닝 워크플로우

색인 단계는 형태소 분석 정보들을 추출하여 데이터베이스에 저장하고 신속하게 검색할 수 있도록 인덱싱 처리한다. 데이터베이스 용량을 줄이고 검색 속도를 높이기 위하여 문화요소 추출에 직접적인 활용도가 낮은 기능어(functional word)들은 삭제된다. 본 시스템의 활용도를 높이기 위해 원 문장의 용례를 복원할 수 있도록 데이터베이스 구조를 설계하였다.

검색 단계에서는 형태소들의 공기(co-occurrence) 정보에 기초하여 질의 형태소의 문맥 형태소들을 문화 요소 후보로 간주하고 이들 후보들의 가중치를 계산하여 가중치가 높은 문맥 형태소들을 최종 문화요소로 추출한다. 본고의 텍스트마이닝 과정에서 사용한 질의(query) 형태소인 동사 '먹다'의 경우, 동사 '먹다'와 한 문장에서 같이 사용된 모든 형태소가 문화요소 후보가 된다. 각 후보들에 대해 공기 빈도를 기반으로 가중치를 계산하고 랭킹함으로써 최종적으로 문화요소를 추출하는 것이다.

### 3. ingestion 프레임과 프레임 구성요소(Frame\_element)

틀의미론에서 말하는 프레임(frame)은 단어의 의미 파악을 위하여 필요로 하는 백과사전적 지식들의 하나로, 어떤 단어의 의미 파악을 위해서는 이와 관련된 지식 체계 전체에 대한 이해가 선행되어야 하고, 이와 관련된 연관 어휘는 동일 프레임의 어휘 집합(lexical set)이 되는 것이다. 틀의미론에 기반한 프레임넷(FrameNet) 프로젝트가 미국의 버클리 대학을 중심으로 시작되어 현재에는 중국어, 일본어 등으로 까지 확장되고 있는데, 본고에서는 프레임넷에서 언어보편론에 기반하여 정의된 다음 ingestion 프레임의 핵심(core) 프레임 구성요소(FE)인 도구(instrument)와 문화적으로 밀접한 관계를 맺고 있는 음식(ingestibles)을 추출하여 한국어와 일본어의 상호 문화권에서 섭취되는 음식물의 특성, 유행, 상호문화의 유입양상 및 영향관계를 고찰한다.

**Ingestion** [Lexical Unit Index](#)

**Definition:**  
 An **agent** consumes food or drink, **ingestibles**, which entails putting the **ingestible** in the mouth for delivery to the digestive system. This may include the use of an **instrument**. Sentences that describe the provision of food to others are NOT included in this frame.  
 The **subject** **DEVOURS** the **carcass** **voraciously**.

**FEs:**

**Core:**

- Ingestibles (Ingestible)**: The **ingestible** are the entities that are being consumed by the **agent**.
- Agent (Agent)**: The **agent** is the person eating or drinking.  
 Semantic Type: **Semantic Type: Agent**

**Not-Core:**

- Extent (Extent)**: The extent to which the **ingestible** are consumed by the **agent**.  
 The **subject** **DEVOURS** the **carcass** **voraciously**.  
 Semantic Type: **Degree**
- Duration (Dur)**: The length of time spent on the ingestion activity.  
 They've been **EATING** **for hours**!
- Instrument (Inst)**: The **instrument** with which an intentional act is performed.  
 Semantic Type: **Physical\_entity**  
**Wearing** **goggles** of performing an action.  
 Semantic Type: **Manner**
- Means (Mean)**: An act performed by the **agent** that enables them to accomplish the whole act of ingestion.  
 The thing **ATE** **by sucking his long tongue out and rubbing with it**.
- Location (Loc)**: Where the event takes place.  
 Semantic Type: **Location**
- Anticipation (Ant)**: The action that the **agent** hopes to bring about by ingesting.  
 Semantic Type: **State\_of\_affairs**
- Source (Src)**: Place from which the **agent** takes the **ingestible**.  
 Semantic Type: **Source**
- Time (Time)**: When the event occurs.  
 Semantic Type: **Time**

<https://framenet.icsi.berkeley.edu>

〈그림 3〉 Ingestion\_frame 정의 및 프레임 구성요소

### 3.1 한·일 ingestibles의 유사성

지리적으로도 역사적으로 상호 문화 교류가 빈번할 수밖에 없었던 한국과 일본은 일상생활에서 빼놓을 수 없는 음식문화에 있어서 서로 영향을 주고받아 많은 유사성을 발견할 수 있다.

〈표 1〉 한·일 ingestibles의 상위랭킹 비교

순 위	한국어 ingestibles	t-score	일본어 ingestibles	t-score
1	밥	31.535965	御飯	35.72893
2	고기	13.113911	肉	21.760323
3	빵	11.325254	パン	21.036144
4	떡	11.018668	ラーメン	20.175919
5	라면	7.8325915	汁	12.914986
6	죽	7.6222463	餅	11.976817
7	국수	7.0906296	麵	11.655988
8	국	6.430256	粥	11.184735
9	보리밥	5.2159863	御握り	9.872373
10	돼지고기	5.1545286	牛肉	8.613544
11	쇠고기	5.1238794	豚肉	7.45127

각 ingestion 프레임의 주요 프레임 구성요소인 ingestibles의 주요 상위랭킹을 살펴보면, 양국 모두 '쌀'이 최상위의 위치를 찾아하는 것을 알 수 있다(밥: 31.535965, 御飯: 21.760323). 이는 공통적으로 주식을 밥으로 하는 쌀문화권에 속한 양국의 특징이 나타난 것이라고 할 수 있는데, 한국어가 '떡: 11.018668', '죽: 7.6222463', '국: 6.430256', '보리밥: 5.2159863', 일본어가 '粥: 11.184735', '餅: 11.976817', '汁: 12.914986', '御握り: 9.872373' 등 쌀과 관련된 ingestibles가 양국 모두 상위랭킹을 점유하고 있다.

일반적으로 일본의 식문화가 한국의 식문화와 뚜렷하게 다른 점으로, 일본에서는 쇠고기, 돼지고기, 닭고기 등의 '육조리문화'가 발달하지 못했다는 점으로 지적된다.(권순자 2005, 7p) 그러나, 실제 분석결과에서는 두 문화권 모두 쌀 다음으로 '고기: 13.113911', '肉: 21.760323'가 각각 차위 랭킹을 차지하는 것으로 나타났는데, 이는 분석대상 코퍼스가 비교적 최신의 것임에 기인하여 양국 모두 근래에 들어 육류 음식이 크게 유행한 것임을 알 수 있다.

〈표 2〉 고기 소비량 변화(한국, 1955-2008)

year	meat consumption		beef		pork	
	total	per person	total	per person	total	per person
1955	41,517	▲▲▲	11,043	-	24,354	-
1960	89,043	3,568	12,950	519	58,025	2,325
1965	97,600	3,401	27,261	950	55,881	1,947
1970	165,063	5,251	37,340	1,188	82,546	2,626
1975	224,734	6,370	70,292	1,992	98,848	2,802
1980	432,682	11,349	99,974	2,622	241,842	6,344
1985	592,862	14,398	120,342	2,923	346,274	8,410
1990	853,485	19,909	176,988	4,129	504,799	11,775
1995	1,231,003	27,447	301,217	6,720	661,710	14,750
2000	1,509,587	31,935	402,381	8,512	779,908	16,500
2005	1,512,075	32,144	316,853	6,736	838,479	17,824
2008	1,727,871	35,582	365,116	7,512	926,764	19,100

2012, 이규진·조미숙

〈표 3〉 소고기의 소비량, 생산량 등의 추이(일본, 1980-2010)

	1980	1990	1995	2000	2008	2010
1인당 소비량 (kg/년)	3.5	5.5	7.5	7.6	5.7	-
국내생산량 (만t)	43.1	55.5	59	52.1	51.8	-
수입량 (만t)	17.2	54.9	94.1	105.5	67.1	-
사육마리수 (만마리)	215.7	270.2	296.5	282.3	289	289.2
사육호수 (만호)	36.4	23.2	17	11.7	8	7.4
1호당 사육마리수 (마리)	5.9	11.6	17.5	24.2	35.9	38.9

일본 농림수산성 「축산통계」, 「식량수급표」 (<http://www.maff.go.jp/>(2019/06/20) )

〈표 2〉에서 한국의 경우, 소득 증대와 함께 육류소비가 꾸준히 증대하고 있는 것을 알 수 있는데, 특히 2000년 육류소비를 보면 쇠고기는 1인당 8.5kg, 닭고기는 6.9kg, 돼지고기의 경우는 2001년에 1인당 16.5kg씩을 먹은 것으로 조사되었다. 2000년 국내 돼지고기의 소비량은 수입된 9만 5891t을 포함해 77만 9908t이고, 부위별로 살펴보면 삼겹살이 국민 1인당 3.66kg 꼴인 17만3139t으로 가장 많이 소비되었다. 특히 한국인이 즐겨먹는 삼겹살은 국내 생산량이 부족하여 전체 소비량의 약 3분의 1인 4만 9863t을 수입하였다. 수입돼지고기 부위로는 삼겹살이 비중이 가장 높고 목살, 갈비 등의 순이다(이규진·조미숙, 2012).

일본에서 육류 소비는 고대 불교의 영향과 정치적인 이유에서 금지되었다. 이후 에도시대까지도 이러한 전통이 이어져오다가, 문명개화가 된 메이지유신 때부터 육류를 먹기 시작하여, 근대화가 본격적으로 이루어진 다이쇼 중반에 공업생산액이 농업생산을 앞지르는 산업구조의 변화와 함께 중산계급의 증가하여 다이쇼 데모 크라시라고 불리는 새로운 가치관과 함께 소득수준이 높아진다. 이러한 여파로 서양문화가 보급되면서 일본인의 식사는 서양화되었고, 여기에 중화요리의 유행까지 가세하여 육류음식이 일반화되기 시작하였다. 또한, 2차 세계대전 즈음에 재일교포들의 ‘호르몬 야끼’에서 비롯된 한국의 불고기 문화가 점차적으로 도입되어 현재에는 일본 각지에 ‘야키니쿠 레스토랑’이 번성하고 있다는 것은 주지의 사실이다(권순자, 2005).

한편, 정도의 차이는 있었지만 한국과 일본 모두 2008년 광우병 사태 등으로 쇠고기 소비량이 감소하고 돼지고기 소비량이 증대하고 있는 것을 알 수 있다(표 1), 돼지고기: 5.1545286, 豚肉: 7.45127).

### 3.2 한·일 ingestibles의 차별성

이준서(2016)에서 ingestion 프레임에 있어서 주요 프레임 구성요소(core FE)로 한국어가 숟가락을 일본어가 젓가락을 각각 특징적인 도구(instrument)로 다음과 같은 문화이미지프레임을 제시하였다.



<그림 6> 한국어와 일본어의 'ingestion' 동사의 문화이미지프레임 비교(이준서, 2013.58pp)

구체적으로 상호 고유음식인 ingestibles를 살펴보면(<표 4>), 일본이 생선음식과 면류 음식을 많이 먹는데 반해(상위 15위 중에 5개 이상), 한국은 '짜개: 4.1984377', '설령탕: 3.1769502' 등 탕류 음식을 많이 먹는 것을 알 수 있다.

<표 4> 한·일 고유 ingestibles 상위랭킹 비교)

순위	한국어 ingestibles	t-score	일본어 ingestibles	t-score
1	떡	11.018668	魚	19.553036
2	김치	9.14266	弁当	16.706926
3	반찬	8.620597	饅頭	15.032057
4	죽	7.6222463	味噌	14.240116
5	국수	7.0906296	醬油	12.689989
6	쌀밥	6.5683165	寿司	12.633821
7	개	6.104271	粥	11.184735
8	된장	5.5115814	刺身	10.953249
9	나물	5.085194	蕎麦	10.844867
10	김밥	4.601607	納豆	10.576638
11	마늘	4.5725784	御握り	9.872373
12	짜개	4.1984377	天麩羅	8.871808
13	떡국	3.5765798	鰻	8.613923
14	고추장	3.2749329	団子	7.8974037
15	설령탕	3.1769502	梅干し	7.5127306

한국이 떡, 쌀밥, 김밥, 떡국, 설령탕 등 쌀을 주식으로 하고 김치, 반찬, 나물, 짜개, 마늘 등이 이러한 주식을 보조하는 역할을 하는데 반해, 사면이 바다로 둘러싸인 일본이 魚, 寿司, 刺身, 鰻 등이 주를 이루고

醬油、餛飩 등이 이를 보조하는 것을 알 수 있는데, 이러한 특징은 ingestion 프레임의 주요 구성요소인 instrument의 대비적인 결과(한국어 ‘숟가락’, 중국어 ‘젓가락’)와 서로 호응하고 있다고 할 수 있는 것이다.

### 3.3 한·일 유행음식

한국어의 ‘빵’이라는 단어의 어원은 포르투갈어이다. 우리나라에서 빵이라는 말을 쓰게 된 것은 우리보다 먼저 서양과 접촉했던 일본의 영향 때문이다. 일본은 포르투갈어의 ‘팡’(pão)이라는 단어를 받아들여 팡(パン)이라고 했는데, 이것을 우리나라도 그대로 받아들였다.

우리보다 훨씬 전부터 적극적으로 개방적인 자세로 외래문화를 수용해온 일본으로부터 외국의 음식문화가 한국에 전파된 것은 1876년(고종 13년) 강화도조약으로 이를 계기로 한반도에 대거 유입되기 시작한다. 비록 한반도 전역에 영향을 미치지 않는 않았지만, 일본음식뿐만 아니라 일본을 통해 서양음식, 중국음식 등이 한반도에 유입되기 시작했다. 이때 자리 잡은 일본 음식이 오늘날까지도 한국음식에 큰 영향을 미치고 있는 것이다. 일본 간장을 베이스로 하여 무, 가마보코, 곤약, 삶은 계란 등을 넣은 국물요리를 의미하는 ‘오뎅:2.6209688’ 이 해방 이후 어묵이 되었으며, 서양식에 일본식을 가미한 와요카시는 호두과자의 탄생에 영향을 미쳤다. 특히 생선과 관련된 일본 음식이 현재까지도 우리나라의 음식에 많은 영향을 미쳤다고 할 수 있다.

〈표 5〉 한·일 상호 유래 ingestibles 상위랭킹 비교

순 위	한국어	t-score	일본어	t-score
	ingestibles (일본 유래 음식)		ingestibles (한국 유래 음식)	
1	생선	6.2958293	焼き肉	11.196357
2	도시락	6.016116	キムチ	7.9155407
3	겨자	5.873839	内臓	5.409124
4	오뎅	3.3615766	胃袋	5.3942876
5	우동	2.6209688	唐辛子	5.140734
6	초밥	2.423969	冷麵	4.7812133
7	생선회	2.3171055	春雨	3.5417757
8	카레	2.2792816	カルビ	3.4063861
9	튀김	2.2313688	レバー	2.9031315
10	계	2.062114	ヒビンバ	2.8772478

근래에 들어 일본에서의 한국 음식의 영향 또한 무시할 수 없다. 특히 2차 세계대전 즈음에 재일교포들의 ‘호르몬 야끼’에서 비롯된 야키니쿠(焼き肉: 11.196357) 등 육고기류의 일본에서의 유행은 최근 들어 한국을 압도할 정도라고 할 수 있다(〈표 1〉, ‘고기: 13.113911’, ‘肉: 21.760323’).

한편, 한국과 일본의 식문화는 글로벌화의 영향에 크게 노출되어있다.

〈표 6〉 한·일 외래 ingestibles 상위랭킹 비교

순 위	한국어 ingestibles (외래음식)	t-score	일본어 ingestibles (외래음식)	t-score
1	아이스크림	5.976276	ラーメン	20.175919
2	겨자	5.873839	カレー	18.496449
3	치즈	4.957618	ケーキ	15.728549
4	햄버거	4.700433	スープ	12.654092
5	샌드위치	4.4191155	サラダ	11.8892145
6	회	4.016477	焼き肉	11.196357
7	샐러드	3.7027493	チーズ	10.622552
8	피자	3.6371226	ステーキ	9.186167
9	스테이크	3.5105488	ピザ	9.14182
10	스파게티	3.2947125	ギョーザ	8.901802

맥도널드와 KFC, 피자헛 등 미국에서 체인화된 외식산업의 다국적 전개는 양국의 식문화에 큰 영향을 끼치고 있다. 식문화에 국한되지 않고, '맥도날드화(McDonaldization, Ritzer, 1996:18)'라고 불리는 규격화 표준화가 세계적으로 퍼져가고 있는데, 이는 양국의 라이프스타일에도 지대한 영향을 미치고 있는 것이다.

#### 4. 맺음말

본고는 데이터마이닝의 일종인 텍스트마이닝 기법을 활용하여 한국어와 일본어의 대응량 코퍼스 기반 한·일문화요소추출시스템으로 'ingestion 프레임'의 주요 프레임 구성요소인 'ingestibles'을 추출하여 상호 대조·비교한 논문이다. 지리적·역사적으로 인접하며 언어적으로도 유사한 동일 한자문화권에 속하여 쌀을 주식으로 삼아온 한국과 일본의 식문화에서 유사한 상호 공통분모가 발견될 가능성이 농후할 것으로 예측되지만, 이준서(2013)에서 제시한 'ingestion 프레임'에서 서로 다른 도구(instrument)를 나타낸 것에서 착안한 본 논문의 텍스트마이닝 분석 결과, 다음과 같은 네 가지 점을 알아낼 수 있었다.

1. 양국 모두 '밥', 'ご飯'이 1위의 조응도를 보이고 있어 공히 '쌀문화권'의 전통을 유지하고 있다고 할 수 있다.
2. 한국어의 상위랭킹에는 '숟가락'을 사용하는 '탕요리'가, 일본어의 상위랭킹에는 '젓가락'을 사용하는 '생선요리'가 많은 것은 이준서(2013)의 'ingestion 문화이미지프레임'의 결과와 서로 호응하는 것이다.
3. 최근 일본어가 한국어 보다 'お肉'의 조응도가 높은 것은 재일교포 및 한류의 영향과 관련지어 생각할 수 있다.
4. 양국 모두 '맥도날드화'라고 불리는 글로벌 규격화와 표준화의 영향을 받고 있다.

**【참고문헌】**

- 김종덕 역(1999). 『맥도날드 그리고 맥도날드화』 시유시
- 권순자(2005) 「육류와 동아시아의 식문화 (한·중·일)」 『동아시아식 생활학회 학술발표대회논문집』 동아시아식생활학회 pp.1-20
- 이규진·조미숙(2012) 「근대 이후 한국 육류 소비량과 소비문화의 변화-쇠고기·돼지고기를 중심으로」 『한국식생활 문화학회지』 27(5) pp.422-433
- 이준서(2013) 「틀의미론(Frame Semantics)에 입각한 일본어 'ingestion' 동사의 '문화 이미지프레임 (Cultural-Image-Frame)' 연구」 『일본연구』 20 pp.51-66
- 이준서(2016) 「다국어 '문화요소추출시스템(CEMS)' 개발 구상」 『일어일문학연구(99)』 pp.289-304
- Fillmore Charles, J.(1975) An alternative to checklist theories of meaning, In: *Proceedings of the First Annual Meeting of the Berkeley Linguistic Society*, Berkeley Linguistic Society, Berkeley, CA, pp.123 - 131
- Fillmore Charles, J.(1977) Scenes-and-frames semantics, *Linguistic Structures Processing*, North-Holland, Amsterdam, pp.55 - 81
- T. Talhelm & X. Zhang.(2014) Large-Scale Psychological Differences Within China Explained by Rice Versus Wheat Agriculture, *Science* 344 pp.603-608
- Un Yong Nahm & Raymond J.(2002) Text Mining with Information Extraction, *SpringSymposium on Mining Answers from Texts and Knowledge Bases*, pp.1-8

〈 要 旨 〉

텍스트마이닝을 활용한 「ingestion 프레임」의 日·韓 「ingestibles」에 관하여 一 考 察

本稿はデータマイニングの一種であるテキストマイニングを活用して韓国語と日本語の大容量コーパスから「ingestion 프레임」のコアフレーム要素の一つである「ingestibles」を抽出して対照・比較したものである。李 (2013) で示した両言語の「instrument」の特徴的な違い(韓国語「スプーン」、日本語「お箸」)から引き出した仮説(「instrument」の違いは「ingestible s」の違いを招き出す)をビッグデータ分析技術であるテキストマイニングから検証しようとした。地理的かつ歴史的に隣接して言語的にも類似した同一漢字文化圏に属して「米」を主食とした韓国と日本の食文化から類似した相互共通分母が見つかる可能性が濃厚であると予測されるものの、データ分析の結果(今回の分析対象のコーパスに限る)、1. 両言語ともに「米」との相応度が一位にランキングされたことから「稲作文化圏」である、2. 日本語のほうが韓国語より「お肉」との相応度が高い、3. 韓国語の上位ランキングには「スプーン」を使う「鍋料理」が、日本語の上位ランキングには「お箸」を使う「魚料理」が多い、4. 両言語ともに「マックドナルド化」といったグローバル規格化や標準化の影響を受けていることが分かったのである。

論文分野：意味論

キーワード：テキストマイニング、データマイニング、フレームネット、文化要素、Ingestion\_フレー

■ 이준서(李俊瑞)

성결대학교 교수

jslee@sungkyul.edu

■ 投稿日 :	2019년	9월	27일
■ 審査開始 :	2019년	10월	19일
■ 審査完了 :	2019년	11월	8일
■ 掲載確定 :	2019년	11월	14일