

The prospect of text-mining:
Semantic network, machine learning, and generative AI

Leo Kim
CEO of Ars Praxia

Ars Praxia

Ars Praxia solves problem by innovative data science.

Ars Praxia delivers strategic insights and interprets the world by actively integrating small & big data



INTRODUCTION



ARS PRAXIA
ANALYZE & CREATE

Art : Sociology

Practice : Technology

*Social
psychology*

Ethnography

LINK

Text
Mining

Semantic
Network

FRAME THEORY, George Lakoff



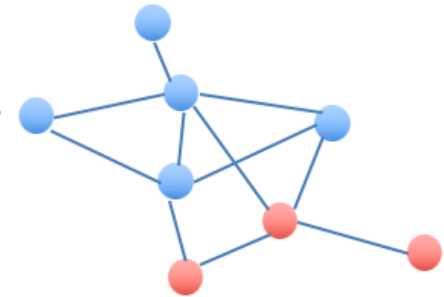
Dr. George Lakoff
University of California at Berkeley

Frame

- Intuitive **structure of linkage** that decides the relation between essence and meaning, event and fact in the process of perceiving political and social agenda
- Frame is reproduced and becomes the way how people perceive the world and develop opinion

Metaphor

- Rhetoric that structures our daily perception
- Cognitive system with **linkage of concepts**
- Potential signifier to study actor's perception, attitude and action



Theoretical legacy and its (re)interpretation

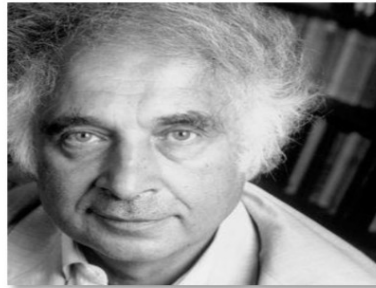
Theory

Structural
Functionalism

Social
Representation

Frame
Theory

Authors



Key concepts

- *Denotation*
- *Connotation*
- *Studium*
- *Punctum*

- Segmentation of social class
- Anchoring
- Objectification

- Political frame
- Cognitive system with linkage of concept/metaphor

(Re)Interpretation

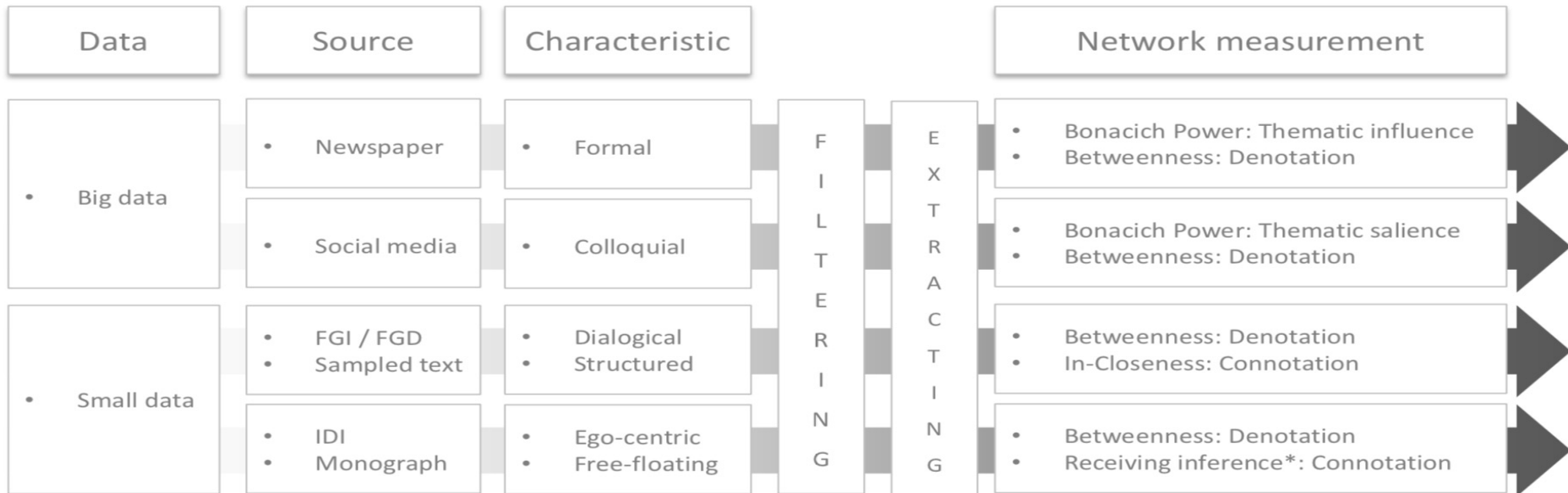
- Denotation as commonly accepted, bridging rhetoric
- Connotation as culturally converging concept

- *Thematic* classification instead of social class
- Denotative anchoring
- Connotative objectification

- A frame reflecting *form of living* (Wittgenstein)
- Measurable and classifiable pattern of linkage

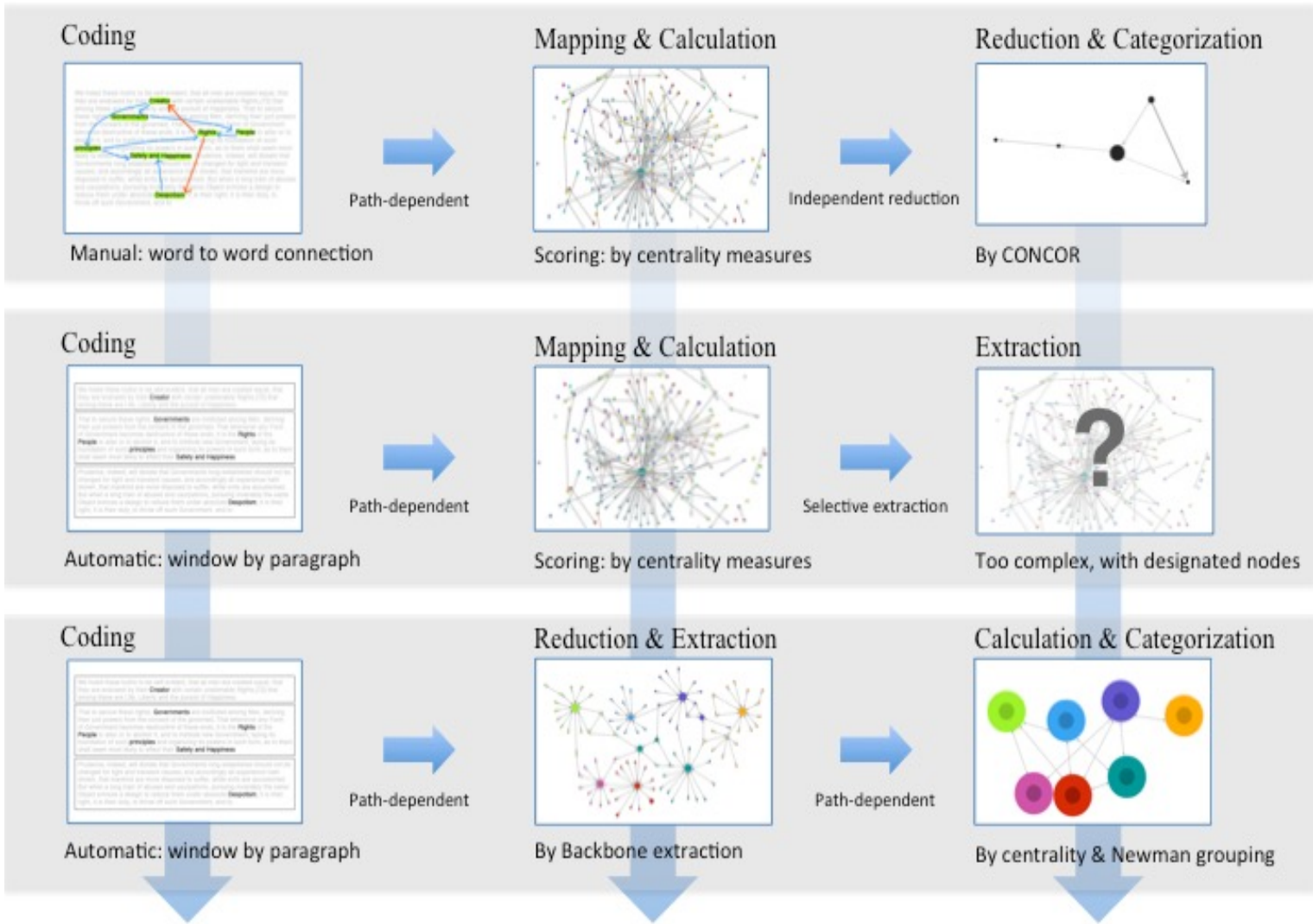
Development of new semantic network measures

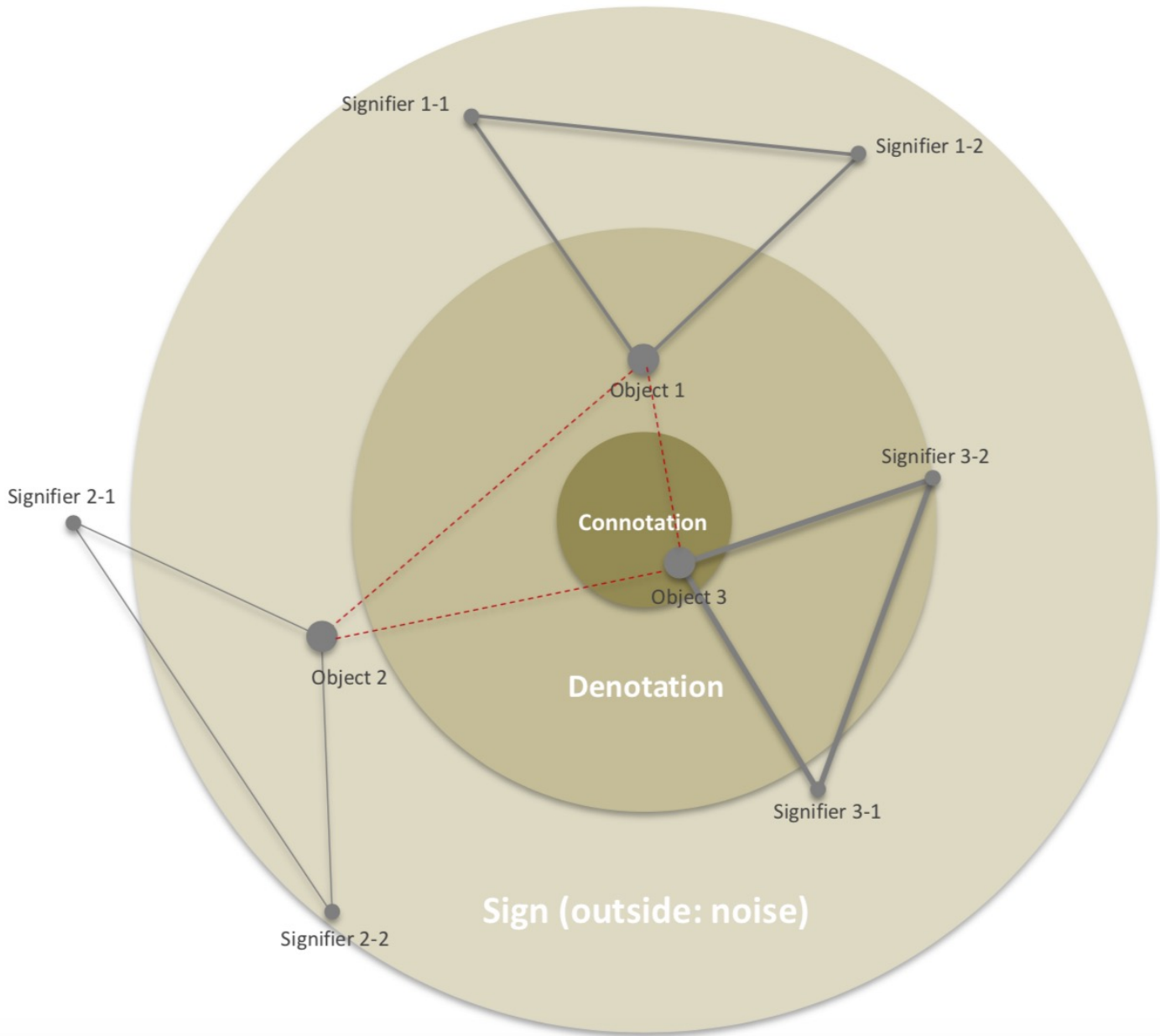
- Existing SNA measures are not always adequate for semantic understanding of network
- Different measure should be developed and applied according to different type of data set



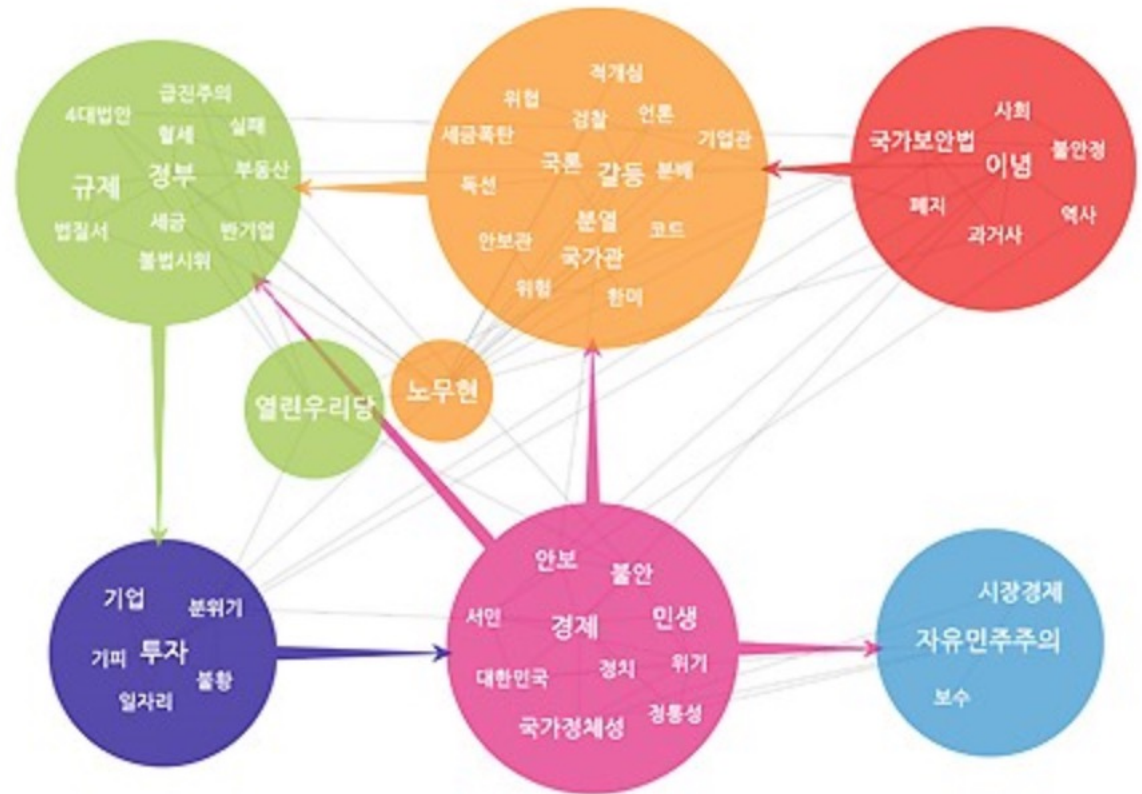
* Receiving Inference finds the pseudo-probability that starts from a random word and converges to a certain word

$$R.Infer = OccurProb \times (I - A)^{-1}$$





THE BEGINNING: What "trust" means for president Park?





네트워크 분석 통해 시대정신 포착한다



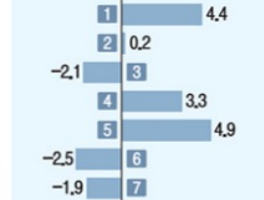
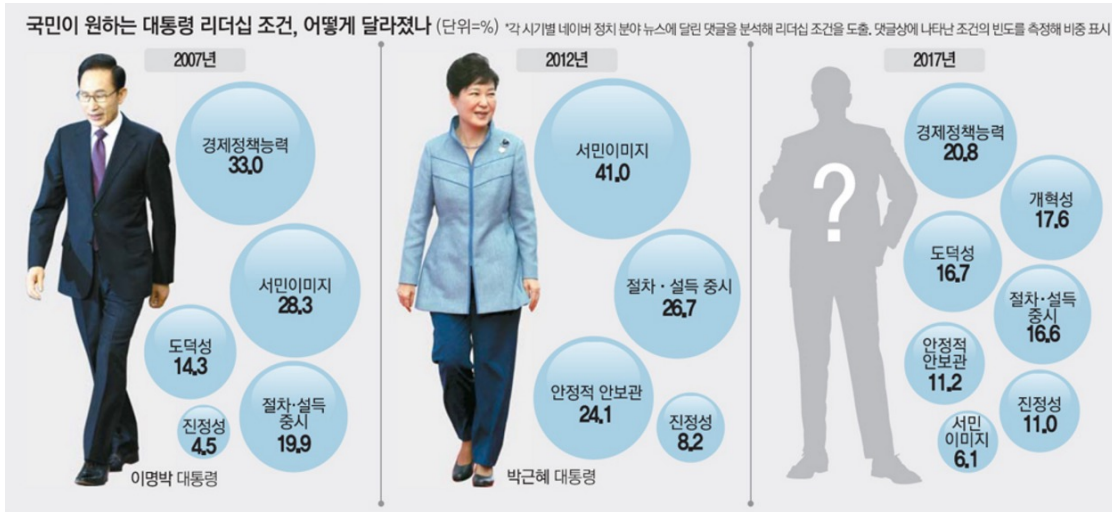
<시사IN>은 최근 사회조사 방법론으로 급부상하고 있는 ‘네트워크 분석’을 도입해 2012년 민심을 탐구하는 연중 기획을 준비했다. 소셜 네트워크 분석 전문기업 트리움(TREUM)과 공동으로 준비한 이 기획을 통해 △정치 소비자인 대중의 사고와 욕망을 심층적으로 파헤치고 △이에 부응하려는 정치 공급자인 정당과 정치인의 시도가 어떻게 성공하고 왜 실패하는지를 분석한다. 이를 위해 <시사IN>과 트리움은 담론 네트워크 분석, 포커스 그룹 인터뷰와 담론 분석의 결합, SNS 분석, 인맥지도 분석 등 매달 새로운 주제와 접근법을 동원해 한국 사회 여론 지형과 정치 지형을 다각도로 분석할 계획이다.

분석을 총괄할 트리움 김도훈 대표(사진)는 “1년간의 분석 작업을 통해 2012년의 진짜 시대정신을 포착하는 것이 목표다. 정치적 주제뿐 아니라 사회문화적인 분석, 세대 분석도 병행할 계획이다”라고 말했다. 이를 위해 김 대표는 “시대정신을 결정하고 선거 결과를 가르는 중도층·회색 지대·스윙보터(부동표)의 민심 향방은 어디로 갈 것인가, 또 이들을 동원해 승리를 거둬야 하는 진보와 보수는 어떤 전략을 취할 것인가를 <시사IN>과 함께 살펴볼 것이다”라고 말했다.

WITH MBC: Revealing communicative strategy of politicians



WITH 매일경제, MBN: People's changing desire for political leadership



대선 주자별 리더십 조건 적합도 (단위=점)

*가나다순. 적합도는 -10~10점. 각 대선주자를 언급한 댓글을 분석해 항목별 빈도를 측정한 뒤, 통계기법을 이용해 점수화함.

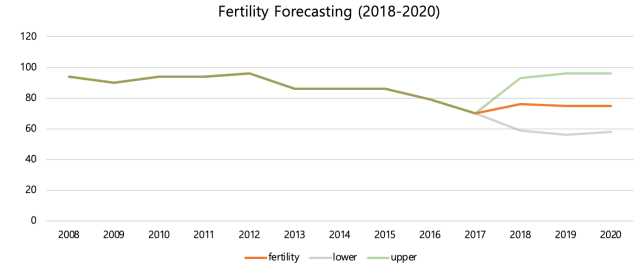


WITH KBS: What the North Korean leader wants

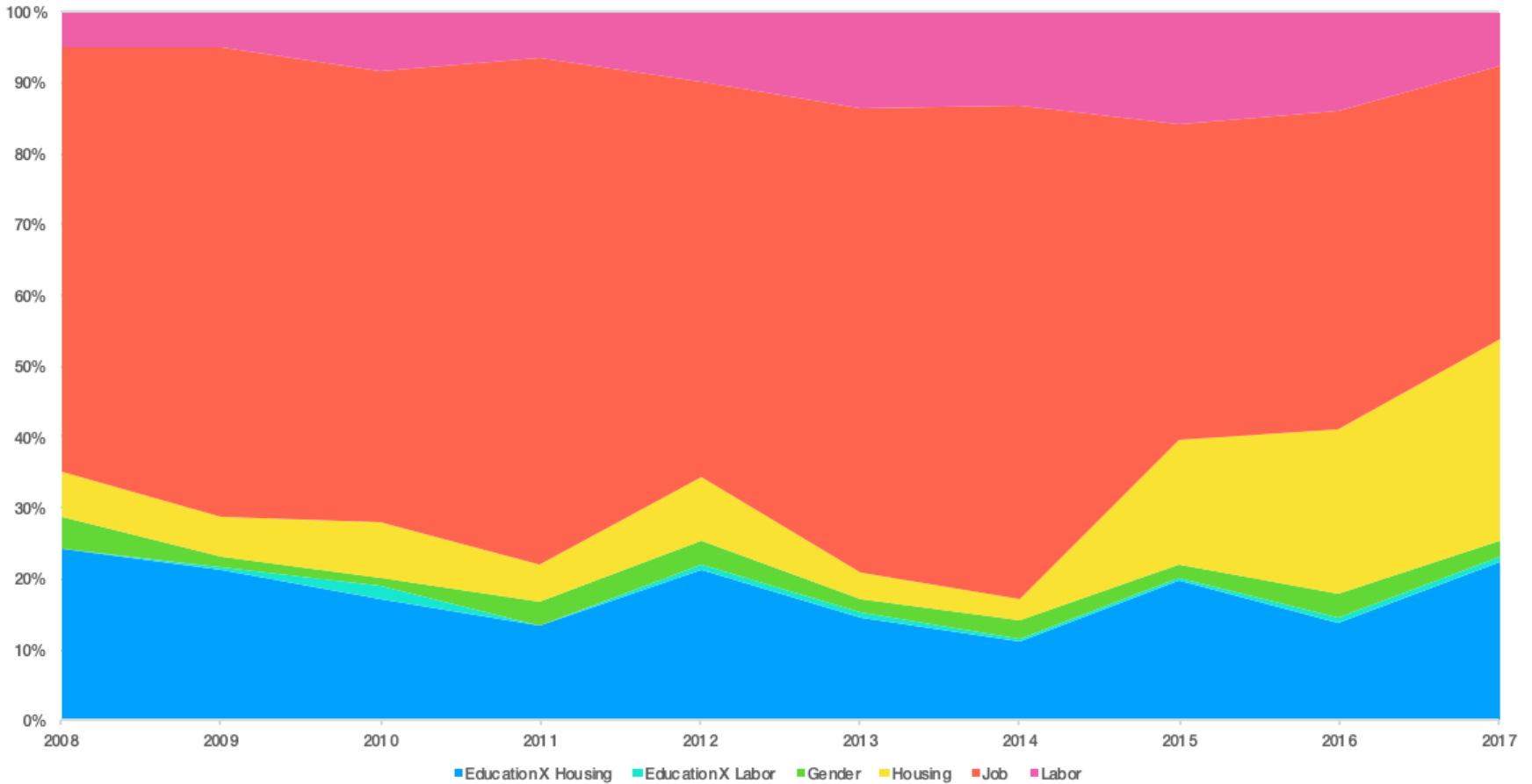


Machine learning (text analysis) and statistical prediction of fertility rate

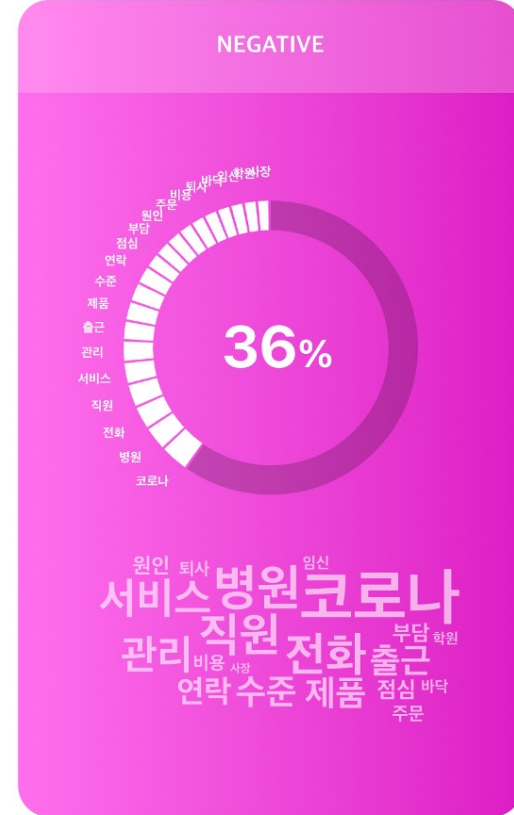
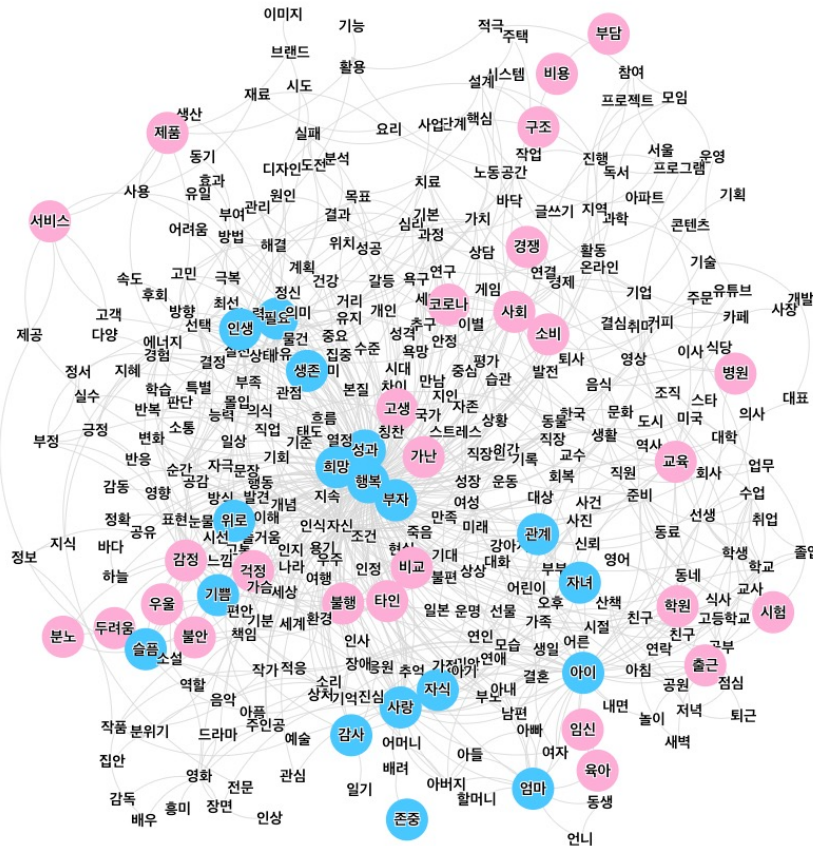
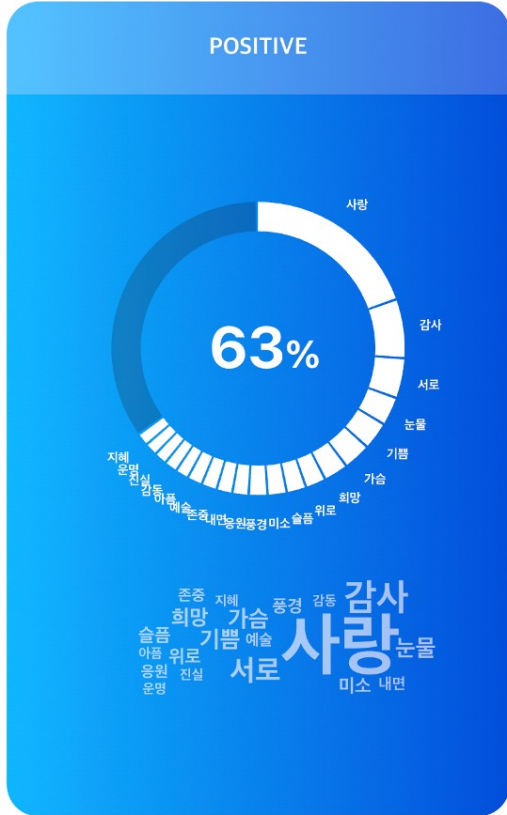
year	fertility	fertility_rate	housing	edu_housing	education	labor	job	employment	income
2008	94	1.192	4.094	5.429	9.455	3.829	6.342	4.159	8.461
2009	90	1.149	3.970	5.328	9.497	3.850	6.469	4.143	8.446
2010	94	1.226	4.419	5.176	9.354	4.443	6.491	4.149	8.474
2011	94	1.244	3.892	4.804	9.228	4.111	6.488	4.157	8.491
2012	96	1.297	4.220	5.050	9.164	4.277	6.016	4.164	8.528
2013	86	1.187	3.638	5.063	9.052	4.977	6.560	4.168	8.535
2014	86	1.205	3.296	4.564	9.149	4.710	6.380	4.184	8.556
2015	86	1.239	4.963	5.056	9.271	4.836	5.878	4.188	8.565
2016	79	1.172	5.670	5.142	9.129	5.153	6.317	4.191	8.563
2017	70	1.052	6.234	6.006	9.122	4.927	6.547	4.199	8.591



저출산 요인별 비중

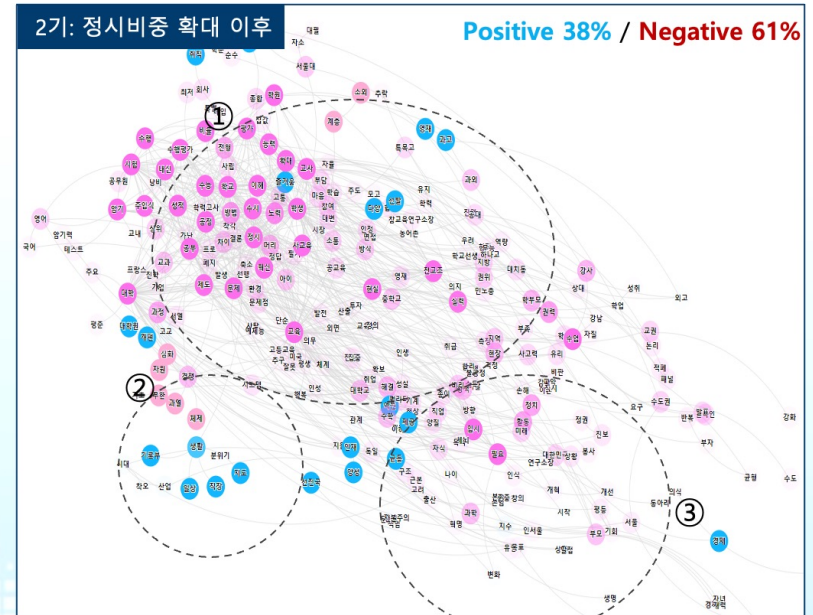
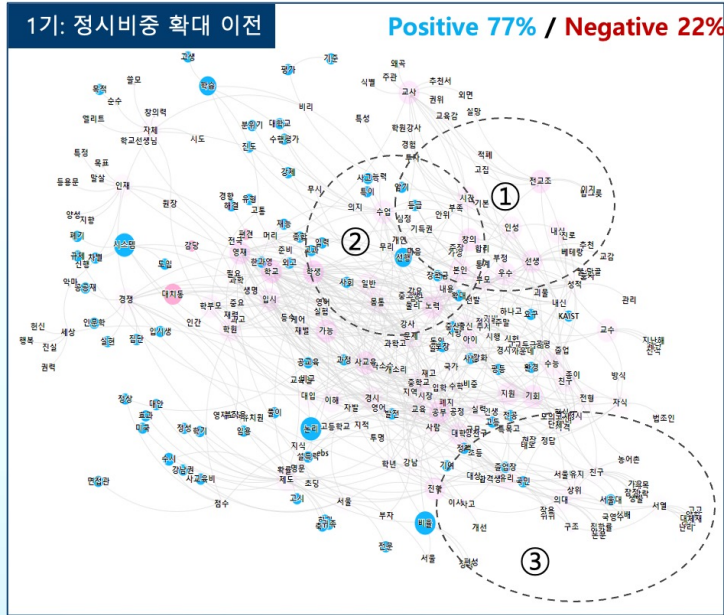


Why Koreans are unhappy



Assessment of national agenda with unstructured data

- 1기에는 ①창의성(인성)이 육성되지 못하는 입시체계, ②선행학습으로 만들어지는 영재교육 비판, ③사고력, 사회 기여를 증진하는 인재 양성 방향이 언급됨
- 2기에는 ①입시 위주 접근이 미래역량에 역행한다는 비판, ②대학원 중심 교육과정 개편과 무한과열경쟁 완화, ③장기적 관점의 인재양성 필요성이 논의됨



2021 10대 국가의제 (KIEP)

- 기술혁신과 신산업육
- 미래인재 양성
- 저출산 고령화
- 코로나 19
- 기후변화 대응
- 사회양극화 해소
- 국가균형발전
- 사회적 대화 및 국민통합
- 남북관계와 한반도평화
- 국제협력과 미-중 갈등 대응

2022 10대 국가의제 (KDI)

- 대전환을 위한 거시경제운용
- 기술패권 경쟁시대 혁신생태계
- 탄소 중립과 저탄소 산업 이행
- 공정사회
- 청년 미래역량 강화
- 코로나19 양극화 극복
- 평생학습사회 구현
- 지방대학 위기와 교육 개혁
- 문화강국을 향한 지역문화 창달
- 분권형 지역발전 행정체계 개편

분석 카테고리

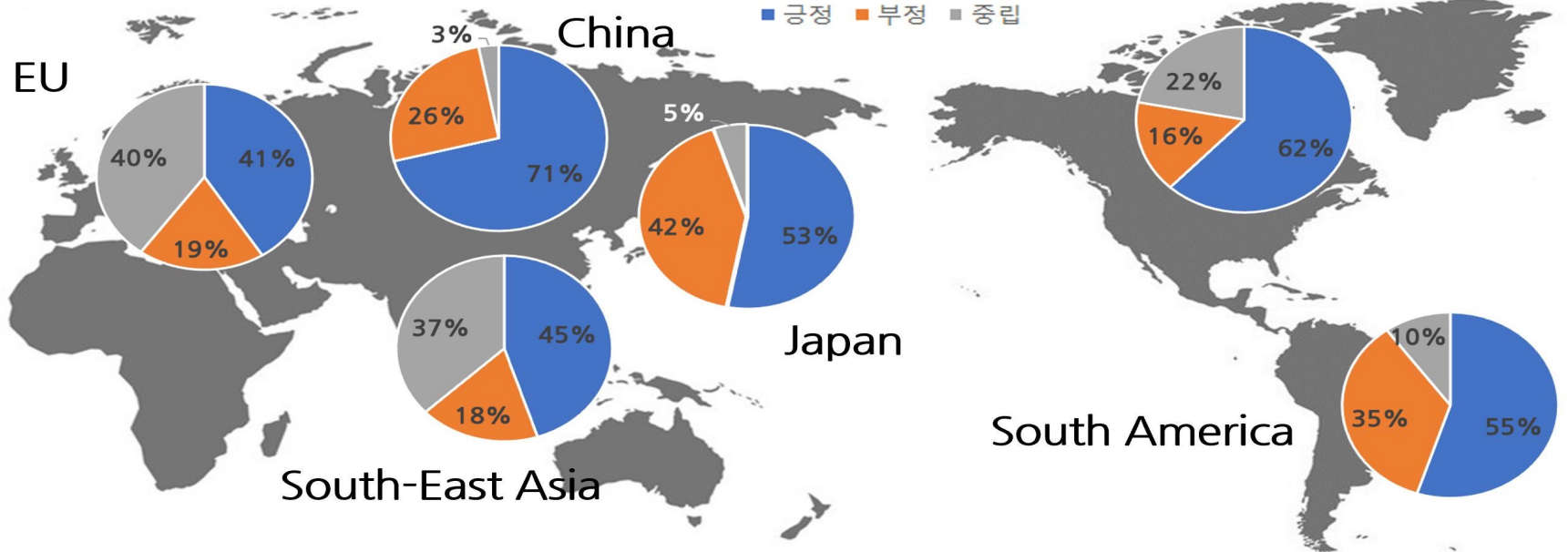
- 거시경제
- 혁신 생태계
- 탄소 중립
- 공정 사회
- 양극화 문제
- 미래역량교육
- 지방균형발전

세부 주제

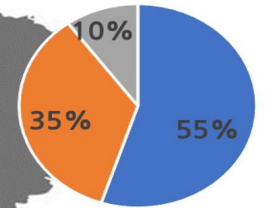
- 대전환 시기의 거시경제운용
- 기술패권 경쟁시대 혁신생태계 조성
- 탄소 중립과 저탄소 산업구조로의 이행
- 공정사회에 대한 프레임과 사회적 대안
- 코로나19 경제·사회 양극화 극복
- 청년미래역량, 평생학습사회 구현, 교육 개혁
- 지방대학위기, 지역문화 창달, 행정체계 개편

Generative AI & cultural understanding

North America



South America



EU

긍정단어	부정단어
Love	Bad
Good	Stop
Thank	Problem

긍정단어	부정단어
可爱(귀엽다)	允悲(슬프다)
开心(즐겁다)	可伶(불쌍하다)
魅力(매력)	失望(실망하다)

China

긍정단어	부정단어
好き(좋다)	反日(반일)
可愛い(귀엽다)	嫌い(싫다)
面白い(재미있다)	悪い(나쁘다)

Japan

긍정단어	부정단어
Like	Sad
Happy	Wrong
Best	Dead

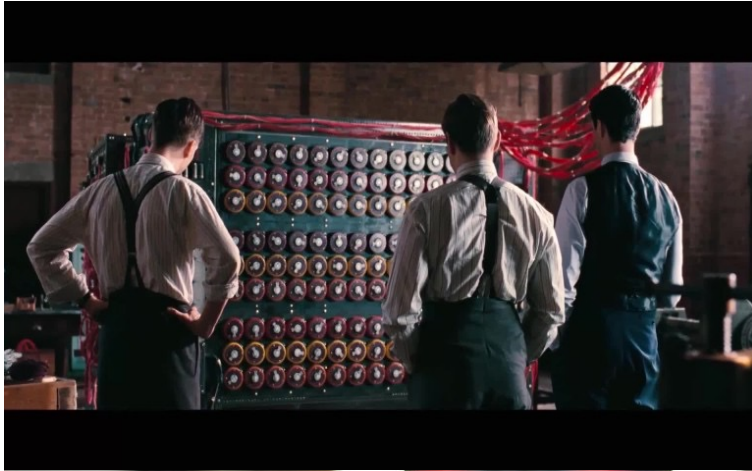
South-East Asia

긍정단어	부정단어
Amazing	Hate
Great	Shit
Favorite	Bias

긍정단어	부정단어
Beautiful	Apologize
Respect	Racism
Truth	Rumors

South America

Future implication



Thank you for listening