

빅데이터의 정확한 언어와 모호한 언어

김형석

한국과학기술원(KAIST) 경영대학 교수(경제학 전공)

전주곡: ‘빅데이터’는 시대정신(Zeitgeist)인가?

바야흐로 ‘빅데이터’ 시대이다. 소위 ‘빅데이터 전문가’라고 불리는 일군의 무리가 마치 중세 시대 복음서를 무시무시한 방식으로 전파하는, 어떠한 토라도 단다면, ‘마녀’로 낙인찍혀 재단의 불쏘시개가 될 것이라고 엄중한 예언을 하는, 구글 신(Google 神)은 알고 있다며 ‘구글 신, 구글 신 구글 신’ 염불하는, ‘구글 신’이 못 박힌 십자가를 앞세워 이교도의 군대로 겁 없이 행군하는 ‘구글교’ 순례자들의 환영을 재현한다는 의미에서의 ‘빅데이터’ 시대이다. 좀 더 직접적으로 표현한다면 전쟁터이건 유대인 수용소이건 베를린의 음악당에 서건 ‘바그너(Wagner)’를 합창하는 나치의 순례자들처럼, 국어학자의 학술 지이건, 회장님의 집무실이건, 회장 사모님들의 미술관 ‘포럼’에서건, ‘빅데이터’를 화학조미료 뿌리듯이 뿌리는 ‘포스트모던’한 ‘셰프’의 감각과도 같다는 의미에서의 ‘빅데이터’ 시대이다. 그나마 유대인의 목숨이 아니라 기업가의 돈주머니와 학생들의 돈주머니를 노린다는 점에서, ‘스왈티카’와 히틀러의 《나의 투쟁》이 아니라 미국 명문 M대학의 권위를 업은 명망가들이 작성한 ‘자연(Nature) 방법서’와 ‘과학(Science) 방법서’를 중심으로 대물주의(大物主義) 광기를 수출한다는 점에서, 나치보다는 인본주의적이다. 또한 이 글을

작성하는 필자의 목숨에 지장을 주지 않는다는 점에서 ‘빅데이터’ 시대는 참으로 존경의 예를 갖추 만하다. 그렇다고 1945년 5월 1일 베를린 함락전의 나치처럼, 아주 미악한 존재라고 보기는 어렵다.¹⁾ ‘구글교’의 순례자들이 ‘극악 무도’하게 불쑥 내민 ‘빅데이터’라는 십자가 앞에서 ‘뱀파이어’라고 낙인찍힐 수 있다는 두려움을, 적어도 인문학자는 체감할 수 있기 때문이다. 그 두려움의 실체를 정확히 밝힘으로써 인문학자가 ‘뱀파이어’가 아님을 소명(疏明)하는 의미에서 아마도 필자에게 이 글의 의뢰가 왔을지도 모르겠다. 의뢰자의 ‘교환 동기’의 가치를 중시하는 경제학자답게, 이 글에서는 ‘빅데이터’의 실체를 해부할 뿐만 아니라, 인문학자의 온전한 두려움 또한 해체하는 것을 목표로 한다.

‘빅데이터’의 정의(定義)

인문학 우상 숭배 타파를 외치는 ‘빅데이터’의 홍위병들은 여러분의 돈주머니와 영혼주머니를 노리지만, 마오쩌둥(毛澤東)의 홍위병들이 그랬듯이, 그들이 외치는 구호 ‘빅데이터’ 자체의 의미는 가늠하지 못한다. ‘빅데이터’란 단어 자체가 주는 ‘사전적’ 의미는 단순히 말해 “데이터가 크구나!”라고 외치는 감탄문의 범주를 벗어나지 못한다. ‘남모를 열등감의 원인’이 ‘크기’에서 비롯될 때, “so big!”이라는 외침은 홍위병의 집단 의식을 일깨우기 마련이다.²⁾ 따라서 전통적으로 왜소 콤플렉스의 악몽을 겪어 왔던 한국적 상황에서, 홍위병들이 외치는 ‘빅데이터’는 자생적인 재구조화 없이, 대물주의(大物主義)와의 야합(野合)을 통하여 시대의 총아, 새로운 패러다임의 재래로 위치가 격상되었다.

1) 베를린의 공식 함락일은 1945년 5월 2일이다.

2) ‘남모를 열등감의 원인’이라는 표현은 한강의 《채식주의자》(창비, 2007)에서 발췌하였다.

그렇다면, 실제 ‘빅데이터’란 무엇을 의미하는가? ‘빅데이터’의 사전적 의미는 정보 수집 기법의 발전에 따른 관찰된 정보량의 비약적 증가, 그렇게 증가된 정보량의 저장 기술 확대를 의미한다. 특히, 여기에서의 정보량은 ‘디지털’화한 정보량, 《표준국어대사전》의 순화된 말로 표현한다면, 0과 1 두 개의 수치로 ‘환원’된 ‘수치적’ 정보량을 말한다. 다시 말하면, 관찰 횟수와 저장 규모 두 개의 차원에서 보았을 때, ‘빅데이터’라 불리는 정보량의 물리적 크기는 말 그대로 “so big!”이다. 예를 들어, 한 사람의 사회적 경험에 대한 정보량을 얻기 위해, ‘빅데이터’ 연구에서는 보통 한 사람에 관한 ‘사회적 행동 변수’ 30개를 정하고, 그 30개 변수를 6분마다 18개월 동안 관찰한다.³⁾ 즉 한 사람에게서 얻는 정보량 또는 데이터는 대략 ‘ $30 \times 240 \times 30 \times 18 = 3,888,000$ ’이다. 주의해야 할 것은 이 데이터 수는 한 사람에게서만 얻어낸 것이라는 점이다. 그러나 ‘빅데이터’를 정의할 때 무엇보다 주목해야 할 것은, 그 정보량이 무엇에 관한 정보량인가? 즉 데이터가 지향하고 있는 지점이 어디인가를 파악하는 것이다. 사실 3,888,000이라는 숫자가 원자 또는 소립자에 대한 데이터 숫자라면 “so big!”이라는 감탄문은 ‘위선적’인 감탄에 가깝다. “외계인이 크다.”라고 말하는 것은, 큰 것은 알지만 우리의 크기와 상관없기에 냉소적으로 말하는 “커요.”에 가깝다. 우리의 감탄이 머무는 지점이 바로 인간의 삶에 관한 정보를 담고 있기에, 우리는 비로소 위선의 가면을 내려놓고 “정말로 커요.”라고 말한다. 다시 말하면, ‘빅데이터’는 인간의 삶에 대한 수치화된 정보량이다. 인간의 삶에 관한 수치화된 정보량이 크기에, 우리는 이것을 ‘빅데이터’라고 부른다. 또한 30개의 ‘행동 변수’가 선사하는 3,888,000개의 숫자가 한 개인의 삶을 이해할 수 있다는 ‘믿음’도 ‘빅데이터’라는 언어에 반영되어 있다. ‘양으로 승부해서 질로 승화할 수 있다’는 ‘소박한 공대생’의

3) 알렉스 펜틀랜드의 《창조적인 사람들은 어떻게 행동하는가》(박세연 옮김, 와이즈베리, 2014) 참조.

믿음을 ‘빅데이터’라는 ‘시적’ 용어로 표현한 것이다. 결국 정보 기술의 발전에 의한 방대한 양의 정보를 수집하고 저장할 수 있는 물리적 기술력이라는 ‘하부 구조(substructure)’ 위에, ‘인간의 삶조차 양으로 승부하면 그 삶의 질을 이해할 수 있다’는 ‘상부 구조(superstructure)’를 가진 믿음 체계로 볼 수 있겠다.

삶의 이해 방식으로서의 ‘빅데이터’: 공대생의 재도전

일단 믿음 체계가 형성되면, 특히 그 믿음 체계가 사람을 향해 있는 것이라면, 그 믿음 체계 특유의 삶을 이해하는 방식이 드러나기 마련이다. 믿음 체계로서의 ‘빅데이터’ 또한 특유의 이해 방식을 선사하는데, 이 특유의 이해 방식은 삶을 이해하는 전통적인 방식과는 확연한 대척점에 있다.

우리는 우리의 삶을 이야기할 때, ‘모호하다’고 말한다. 국립국어원 《표준국어대사전》에 의하면, ‘모호하다’는 “말이나 태도가 흐리터분하여 분명하지 않다.”로 정의된다. 따라서 우리의 삶이 ‘모호하다’고 규정짓는 것은 우리의 삶이 흐리터분하여 분명하지 않고, 이런 이유로 그러한 삶을 ‘분명한 말’이나 ‘분명한 태도’라는 매개체(媒介體)로 표현한다 한들, ‘모호함’의 무게가 너무나 무겁기에, 그 무거움에 짓눌린 우리의 언어나 태도 또한 흐리터분하고 분명함과는 거리가 멀 수밖에 없다. 결국 모호한 우리의 삶을 분명한 언어·말로 이해한다는 것은 ‘모호함’의 꼬끼리를 ‘분명함’이라는 부처님의 손가락으로 지탱하려는 위태로운 곡예처럼 보인다. 가끔 찾아오는 위태로운 곡예는 우리 삶에 신선한 자극을 주고 생기를 북돋아 주지만, 그러한 곡예가 일상이라면, 그 곡예는 우리 삶의 일부가 될 수 없다. 즉 위태로운 곡예는 예외적일 때만 가치가 있다. 따라서 ‘모호’한 삶을 ‘분명한’ 언어로 이해하는 것은 우리의 삶을 이해하는 전통적인 방식이 아니다. 우리의 삶을 이해하

는 전통적인 언어 방식은 ‘모호한’ 언어이다. ‘모호함’의 무게에 짓눌려 질식 상황을 처절하게 분명한 언어로 고발하기보다, 그 ‘모호함’의 무게를 온전히 받아들이는, 그 처절하게 숨 막히는 상황과 함께 산화하는 ‘메시아’의 메시지와도 같은 ‘모호한’ 언어가 오히려 이율배반적으로 우리의 ‘모호한’ 삶을 ‘분명히’ 전달하는 것이, 삶을 이해하는 전통적인 언어 방식이다. 그래서 우리의 ‘모호한’ 삶을 ‘모호한’ 언어로 이해하는 전통적인 이해 방식의 정점은 바로 ‘시’이다. 김소월의 <산유화>라는 시를 통해 우리가 우리의 모호한 삶을 어떻게 ‘모호한’ 언어로 이해하는지 알아보자.

산에
산에
피는 꽃은
저만치 혼자서 피어 있네.

최근 화제가 되었던 《시를 읽는 그대에게: 공대생의 가슴을 울린 시 강의》(휴머니스트, 2015)에서 정재찬 교수가 지적했듯이, 시어 ‘저만치’가 의미하는 거리는 수치로 치환(置換) 가능한 물리적 거리가 아니다. 이 시의 화자와 꽃 사이의 거리 ‘저만치’를 1미터, 5미터, 10미터, 200미터로 이해하는 것은 삶의 ‘모호함’이라는 무게에 짓눌려 압사해 버린, ‘분명한 언어’란 이름을 가진 사내의, 피에 버무린 살점이 꽃잎처럼 산화하는 것을 보는 것과 같다. 시어 ‘저만치’는 ‘모호함의 천공(天空)’을 완전(完全)히 짚어질 수 있는, 모호한 존재였던 거인족(巨人族) 아틀라스처럼, 모호함의 무게를 원망 없이, 속절없이 감내하는 모호한 언어의 극치를 보여준다. 시어 ‘저만치’의 거리는 그냥 지나치기는 아쉬운 거리, 그러나 내 손길을 뻗친다 한들, 인연의 끈을 잡을 수 없는 거리, 아련히 다가오는, 그러나 이내 시나브로 멀어져 가는 ‘복학생 오빠’와의 거리, 그래서 가슴 설레는, 한(恨)이 이슬처럼 망울지는, ‘마음’의

‘모호함’을 드러내는 거리이다. 모호한 언어아말로 모호한 삶을 ‘분명히’ 전달할 수 있는 언어임을 김소월의 시어 ‘저만치’가 오히려 ‘정확히’ 보여준다. 화자와 꽃 사이에 벌어지는 삶의 모호한 만상(萬象)을 1미터, 5미터, 10미터, 200미터 등과 같은 ‘정확한 언어’로 이해하는 것은 언어의 폭력성 또는 언어술사의 재능 없음을 드러내는 ‘분명한’ 언어 방식이다. 그렇기에 ‘모호한’ 삶을 ‘모호한’ 언어로 ‘정확히’ 전달하는 시라는 이해 방식이야말로 인문학의 정점이자 극치라는 것이 전통적인 견해이다.

삶의 이해 방식으로서의 ‘빅데이터’는 바로 이 전통적인 견해가 선형적으로 배제해 왔던, ‘분명하고 정확한’ 언어 방식으로 화자와 꽃 사이의 거리를 ‘정확한 잣대’로 재고자 하는 시도이다. 화자와 꽃 사이에 벌어지는 삶의 모호한 만상을 1미터, 5미터, 10미터, 200미터 등과 같은 ‘정확한 언어’로 재단하는 것이 삶의 민낯을 드러낼 수 있고, 그러한 몸짓이 의미 있음을 보이고자 하는 시도이다. ‘양으로 승부하여 질로 승화할 수 있음’을 보이고자 하는 공대생의 재도전이기도 하다. 즉 ‘시’가 추구하는 방식과는 달리, ‘정확한’ 언어로 ‘모호한’ 삶을 이해하고자 하는 시도이다. ‘모호한 천공’을 수많은 숫자로 조합하고 빚어낸, 인공족(人工族) 아틀라스로 지탱하고자 하는 시도인 셈이다.

이제 ‘시’의 화자와 꽃 사이의 거리를 어떻게 ‘정확한 언어’, 즉 수치화된 정보량을 활용하여 이해할 수 있는지, ‘모호한 언어’를 대표하는 시어 ‘저만치’를 어떻게 ‘정확한 언어’로 대체할 수 있는지 ‘빅데이터’ 방식으로 알아보자. ‘빅데이터’ 방식으로 시어 ‘저만치’를 ‘명확한 빅데이터 언어’로 대체하기에 앞서, 먼저 일반적 의미에서의 ‘빅데이터 분석’을 소개하고자 한다.

문법적 은유(grammatical metaphor)로서의 ‘빅데이터 분석’

‘빅데이터’ 방식으로 삶을 이해한다는 것은 기본적으로 인간의 삶을 분석(分析)한다는 것이다. 《표준국어대사전》에 의하면, 분석이란 ‘엮혀 있거나 복잡한 것을 풀어서 개별적인 요소나 성질로 나눔’을 뜻한다. 즉 분석이란 분석의 주체가 분석의 대상 또는 객체를 엮혀 있거나 복잡한 그 무엇으로 전제한 후, 대상의 복잡함과 대비되는 ‘단순함’을 표상하는 개별 요소 및 성분으로 분해하여 대상의 실체를 ‘까발리는’ 것을 의미한다. 따라서 ‘빅데이터’가 이해하는 삶은 분석의 대상이자, 엮혀 있고 복잡한 것이다. 인간의 모호한 삶은 ‘모호한’ 시어로만 이해되는 신비로운 천공이 아니라, 단순히 엮혀 있어 복잡한, 그래서 우리가 풀어나 까발려야 하는 ‘실타래’일 뿐이다. 그 ‘실타래’를 풀기 위한 ‘가위’를 성분 또는 요소라 말할 수 있는데, ‘빅데이터 분석’에서는 그 ‘가위’를 ‘행동 변수’라 부른다. 다시 말해서, 인간의 ‘모호한’ 삶을 ‘행동 변수’라는 ‘명확한 언어’로 분해하는 것이다. 결국 ‘빅데이터 분석’이란 ‘빅데이터’의 사전적 의미, 그 사전적 의미에서 파생된 신념 체계, 다시 신념 체계에서 파생한 삶을 이해하는 방식, 그 이해하는 방식이 보편적 분석의 범주와 결합된 것을 의미한다. 즉 ‘빅데이터 분석’이란 할리데이(Michael Halliday)의 언어 이론을 차용하자면, 지금까지 논의한 ‘빅데이터’의 일차적(congruent) 표현을 문법적으로 은유화(grammatical metaphor)한 것이다.

‘행동 변수’란 관찰 가능한 ‘보편적’ 인간의 직접적 행동과 ‘보편적’ 인간의 의도, 선호 등 직접적으로 관찰 불가능한 주관적 요인을 간접적·반복적으로 관찰할 수 있는 신호들(signals)을 의미한다. 반복적으로 관찰 가능하기에 ‘행동 변수’는 사전적 의미에서 ‘빅데이터’화 할 수 있다. 즉 인간의 특정 행동 및 특정 신호를 나타내는 ‘행동 변수’는 ‘수리 언어화’ 또는 ‘수치화’된 후 ‘빅데이터’란 이름의 정보량으로 기억되는 것이다. 더 나아가, 행동 변수와

행동 변수 간에 인과율 또는 인과관계를 확정할 수 있다면, 우리는 ‘행동 변수’를 통해 인간의 특정 행동이 ‘예측 가능하다’고 말할 수 있다.

여기까지 논의한 ‘빅데이터 분석’은 사실 보편적 과학 분석의 속성에서 크게 벗어나지 않는다. 보편적 과학 분석과 빅데이터 분석의 결정적인 차이점은 우리가 관찰한 ‘행동 변수’를 조종하는 ‘우리의 그림자’가 존재한다는 점이다. 다시 말해서, ‘행동 변수’는 ‘보편적’ 인간의 어떤 선택의 결과로 파생된다.

굳이 비유하자면 ‘빅데이터 분석’은 피아니스트의 피아노 연주를 감상하고 비평하는 ‘음악 평론가의 평론(critique)’과 같다. 분석의 주체가 관찰하는 하나의 ‘행동 변수’는 무명의 피아니스트가 이름 모르는 작곡가의 악보에서 반복적으로 뽑아 오는 특정한 음높이에 위치해 있는 ‘도’ 음을 듣는 것과 같다. ‘분명히’ 피아니스트가 ‘미치광이가 아닌 사람’인 것은 알건만, ‘빅데이터 분석가’라는 명함을 가진 음악 평론가는 오직 특정한 높이의 ‘도’ 음만을 관찰하는 것과 같다. 여러 높이의 ‘도’ 음을 반복적으로 듣고, 또한 여러 높이의 ‘미’, ‘솔’ 음을 구별해서 피아니스트가 들려주는 ‘미지(未知)의 곡’과 이를 파악하는 음악 평론가의 작업을 ‘빅데이터 분석’이라 부를 수 있겠다. 관찰된 데이터 또는 수치화된 정보량에서는 ‘행동 변수’로 명명된 ‘도’, ‘미’, ‘솔’, 또는 ‘도’, ‘파’, ‘라’만 기록될 뿐, 그 행동 변수 ‘도’를 어느 시점에서 선택한 피아니스트의 의도, 목적 등은 데이터에 기록되지 않는다. 그러나 데이터에서 기록되지 않는다고 해서, 피아니스트의 존재가 무시될 수는 없다. ‘분명히’ 피아니스트가 ‘행동 변수’로 피아노 건반을 ‘감촉’함으로써 체계적으로 선택하고 있다는 것을 알기 때문이다. 특히 이 피아니스트가 들려주는 곡의 악보를 사전적(事前的, ex ante)으로, 또는 선험적(先驗的, a priori)으로 알지 못할 때, ‘행동 변수’를 조종하는 ‘우리의 그림자’로서의 피아니스트의 존재는 ‘빅데이터 분석’에서 절대 무시될 수 없다.

과학 분석에서의 ‘행동 변수’는 행위의 시작이자 끝이지, 그 행위를 조종하

는 ‘그림자’를 논하지 않는다. ‘원자의 속도’는 바로 관찰 행위의 시작이면서 마지막에 불과하다. 그 ‘원자의 속도’를 조종하는 것이 있다면, 그것은 또 다른 ‘행동 변수’이어야 하지, 행동 변수가 아닌 ‘그림자’일 수는 없다. 칸트의 인과율 형식에 관한 논쟁에서 ‘방사성 원자 라듐 B가 전자를 방출하고 라듐 C로 변하는 과정에는 어떠한 원인도 없다’는 하이젠베르크(Werner Karl Heisenberg)의 주장은 자연 과학 분석과 인간의 삶을 대상으로 삼는 ‘빅데이터 분석’의 본질적 차이를 오히려 극명히 드러낸다. 반면, ‘빅데이터’의 분석 대상이 인간의 삶이기에, ‘행동 변수’ 수립 시 항상 그 ‘행동 변수’를 선택하는, 그러나 ‘행동 변수’는 아닌, 그렇기에 직접적으로 관찰 가능하지 않은, ‘우리의 그림자’이자 피아니스트를 염두에 두어야 한다.

실전 연습: OK 쿠팡의 사례⁴⁾

앞 장에서 ‘빅데이터 분석’의 의미와 ‘모형적 사고’에 기인한 ‘빅데이터 분석’ 자체의 모호함까지 살펴보았다. 이제 ‘모호한 언어’를 대표하는 시어 ‘저만치’를 어떻게 ‘정확한 언어’로 대체 가능한가를 시현해 보고자 한다.

먼저 ‘빅데이터 분석’에서 가장 중요한 시발점은 관찰 가능한 ‘행동 변수’의 존재 여부를 파악하는 것이다. 첫째, ‘맥락’에 맞는 수치화 가능한 ‘행동 변수’를 설정할 수 있어야 하고, 그 ‘행동 변수’에 대응하는 반복 관찰 가능한 데이터가 존재해야 한다. 앞 장에서 논의했듯이, ‘행동 변수’란 반복적으로 관찰 가능한 ‘보편적’ 인간의 직접적 행동과 ‘보편적’ 인간의 의도, 선호 등 직접적으로 관찰 불가능한 주관적 요인을 간접적·반복적으로 관찰할 수 있는 신호들을 포함한다. 따라서 ‘행동 변수’는 ‘보편적’ 인간의 의도, 선호,

4) 이 절의 분석 내용은 크리스티안 루더의 《빅데이터 인간을 해석하다》(이가영 옮김, 다른, 2015)를 참조하였다. 특히 [그림 1], [그림 2], [그림 3]은 해당 문헌에서 발췌하였다.

동기에 의한 제반 행위를 포괄하기 때문에, 김소월의 시 <산유화>에서 꽃과 화자의 심적 거리 ‘저만치’를 ‘행동 변수’화하는 것이 결코 불가능하지 않다. 먼저 ‘저만치’ 거리에 있는 대상 ‘꽃’은 ‘화자’에게 연모(戀慕)의 대상이다. ‘전통적’으로 연모의 대상은 이성 상대이기에, ‘저만치’에 피어 있는 ‘꽃’은 화자인 ‘내’가 연모의 감정을 느끼는 이성 상대로 분석할 수 있다. 연모의 감정이란 결국 수많은 이성 상대 가운데 하나를 선호하는 것이다. 여기서 우리는 ‘보편적 화자인 나’를 분석하는 것이지, 카사노바를 분석하는 것이 아니다. 따라서 ‘보편적 화자인 나’는 수많은 상대에서 오로지 하나의 ‘꽃’을 선호하고 선택할 수 있다는 행동 강령의 ‘제1원리(the first principle)’를 부여할 수 있을 것이다. 만일 ‘수많은 상대’, ‘꽃’, 화자인 ‘나’를 숫자로 표시할 수 있다면, 또한 그러한 숫자를 모을 수 있다면, 우리는 ‘빅데이터 분석’을 통해 성공적으로 ‘저만치’를 명확한 언어로 대체할 수 있을 것이다. 따라서 먼저 ‘수많은 이성 상대’를 수치화하는 것이 가능할지 여부를 따져야 할 것이다. 만일 반대편에 있는 선택 가능한 이성 상대를 숫자로 표현할 수 있다면, ‘내’가 선택하는 ‘꽃’도 선택 가능할 것이고, ‘꽃’ 입장에서 보면, 화자인 ‘나’도 ‘수많은 이성 상대’ 중 하나이므로 수치로 표현 가능할 것이다. 물론 ‘꽃’ 또한 반대편 성(性)을 대표하는 카사노바인 경우는 배제한다. 다시 말하면, ‘모호한’ 언어 ‘저만치’는 ‘명확한’ 언어 ‘선호도’로 대체 가능할 것이다. 물론 ‘저만치’의 상대를 어떤 숫자로 대표해야 할지 결정하는 것은 쉬운 문제가 아닐뿐더러, 설사 대표하는 숫자를 찾았다 한들, 그 숫자를 기록한 데이터를 모으는 것도 쉬운 문제는 아니다. ‘양으로 승부하여 질로 승화’하려는 공대생의 재도전은 여전히 험난한 길인 것처럼 보인다.

그러나 미국의 데이트 사이트 ‘OK 큐피드’ 운영자이자 ‘데이터 과학자’ 크리스티안 루더(Christian Rudder)는 놀라운 발상으로 ‘데이터 과학자’가 ‘양으로 승부하여 질로 승화’하는 돌파구를 찾아냈다. 일단 매년 1,000만 명 이상이 방문하는 사이트를 운영하기 때문에, 적절한 행동 변수만 정해진다

면, 5년간 자료 5,500만 개를 생성하는 것이 가능했다. 또한 남녀 간의데이트를 성사시키는 사이트를 운영하기 때문에 남녀 간 선호에 관한 정보를 비교적 쉽게 수집할 수 있었다. 크리스티안 루더가 고안해 낸 ‘행동 변수’는 아래와 같다.

- 1) 여성은 어떤 나이의 남성을 가장 매력적으로 생각하는가?
- 2) 남성은 어떤 나이의 여성을 가장 매력적으로 생각하는가?

[그림 1]은 OK 쿠파드 자료를 활용하여 조사한 연령별 여성의 남성에 대한 선호 연령을 표로 나타낸 것이다. 예를 들어, 20세인 여성이 가장 매력적으로 느끼는 남성의 연령은 23세인 것을 의미한다. 또는 20세인 여성은 23세인 남성을 ‘선호’한다. 시어 ‘저만치’는 여성의 남성 ‘연령 선호도’로 대체된다. 표에서 보듯이, 21세인 여성은 23세의 남성을, 22세인 여성은 24세의 남성을, 23세의 여성은 25세의 남성을 가장 선호하는 것을 알 수 있다. 시어 ‘저만치’의 거리는 그냥 지나치기는 아쉬운 거리, 그러나 내 손길을 뻗친다 한들, 인연의 끈을 잡을 수 없는 거리, 아련히 다가오는, 그러나 이내 시나브로 멀어져 가는 ‘복학생 오빠’와의 거리, 그래서 가슴 설레는, 한(恨)이 이슬처럼 망울지는, ‘마음’의 ‘모호함’을 드러내는 거리이지만, 이 연령 선호도는 왜 ‘복학생 오빠’와의 거리가 그토록 한이 이슬처럼 망울지는 거리인지, 숫자로 정확히 보여준다. ‘정확한’ 언어로 말하자면, 20세 초반의 여성은 2~3살 많은 남성을 ‘체계적’으로 선호하는 경향이 있기에, ‘복학생 오빠’가 ‘저만치’ 거리에 자리를 차지할 가능성이 매우 높다.

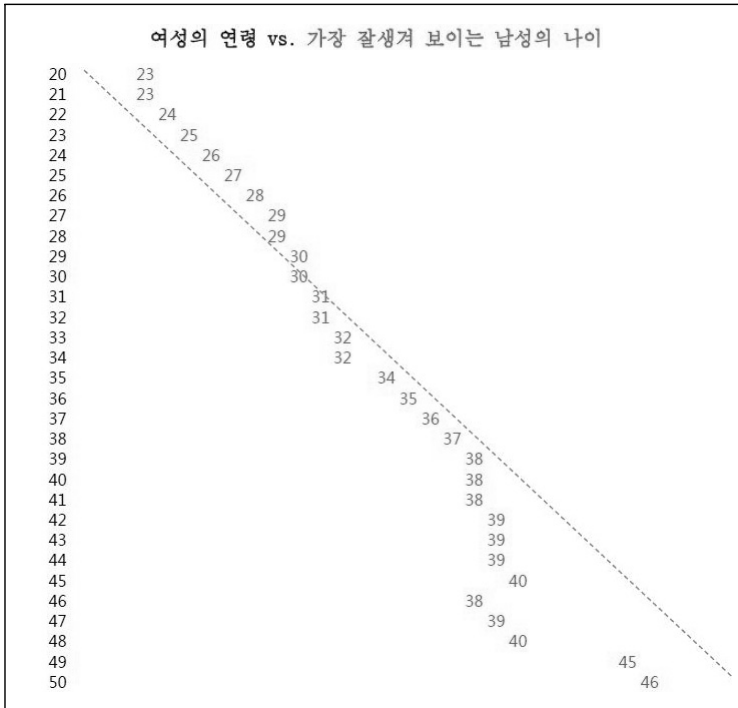
그림 1

여성의 연령 vs. 가장 잘생겨 보이는 남성의 나이	
20	23
21	23
22	24
23	25
24	25
25	26
26	27
27	28
28	29
29	29
30	30
31	31
32	31
33	32
34	32
35	34
36	35
37	36
38	37
39	38
40	38
41	38
42	39
43	39
44	39
45	40
46	38
47	39
48	40
49	45
50	46

또한 [그림 1]은 여성의 나이가 증가함에 따라, 선호하는 남성도 비교적 ‘비례적’으로 올라간다는 것을 알 수 있다. 예를 들어 29세의 여성은 29세 남성을, 30세 여성은 30세 남성을, 31세 여성은 31세 남성을 선호함을 알 수 있다. 그러나 여성의 나이가 32세를 넘어감에 따라, 연하의 남성을 선호하는 경향이 강해지고 있음을 알 수 있다. 놀랍게도, 작금의 ‘송중기’ 열풍 사태를 이 도표는 ‘정확한’ 숫자로 보여 준다. 지금까지의 논의를, 크리스티안 루터가 명명한 ‘나이 비교선’을 가지고 재구성할 수 있다. [그림 2]는

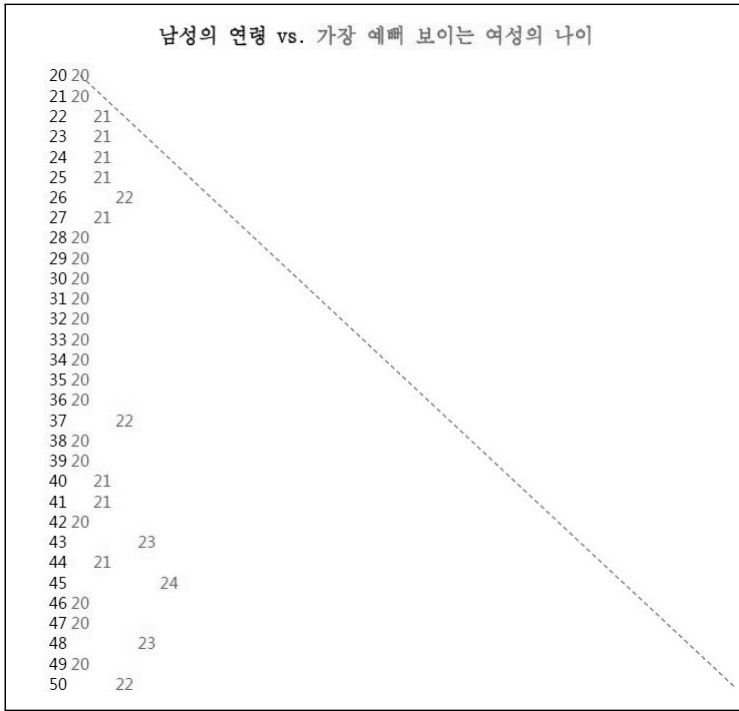
[그림 1]을 여성의 나이와 남성의 나이가 같은 ‘나이 비교선’으로 재구성한 것이다.

그림 2



대각선으로 그려진 ‘나이 비교선’을 보면, ‘복학생 오빠’와의 거리, ‘송중기’ 열풍 사태 등을 보다 명확히 파악할 수 있다. 이제 남성의 연령 선호도를 알아보고자 한다. [그림 3] 역시 OK 쿠피드 자료를 활용해서 작성되었으며, 대각선의 점선은 ‘나이 비교선’을 나타낸 것이다.

그림 3



남성의 경우, '나이 비교선'은 의미가 없다. 남성은 오로지 20대 초반의 여성만이 '저만치' 거리에 피어 있는 '꽃'이기 때문이다. 저만치의 '꽃'이 20대 초반으로 고정되어 있다면, 과연 그 거리를 내 손길이 뻗친다 한들, 인연의 끈을 잡을 수 없는, 아련히 다가오는, 한이 이슬처럼 망울지는 모호하고, 미묘한 거리라고 말할 수 있을까? 남성에게 애당초 같이 늙어가는 원숙미에 대한 존경은 있지도 않다. 남성의 '꽃'에 대한 반응은 지극히 생화학적 반응처럼 보인다. 20세와 21세의 남성은 '저만치' 한이 이슬처럼 망울지는 '시감(詩感)'을 만끽할 수도 있을 것이다. '빅데이터 분석'에 의하면, 20세, 21세의 남성이 한이 맺힐 정도로 아끼는 20세의 여인은 오직 23세 '복학생 오빠' 쪽으로

시선이 고정되어 있다는 것을 ‘빅데이터 분석’이 보여주기 때문이다.

크리스티안 루더가 지적했듯이, 이런 연구 결과는 ‘진실의 이면에 가려져 있는 우리의 허영심과 취약점’을 공공연히 드러낸다. ‘모호한 천공’을 수많은 숫자로 조합하고 빚어낸, 인공족 아틀라스로 지탱했을 때, 우리는 까발려진다. 이런 의미에서 ‘명확한 언어’는 ‘폭력적’이다.

‘빅데이터 분석’의 명과 암: 구글 독감 예측 프로그램

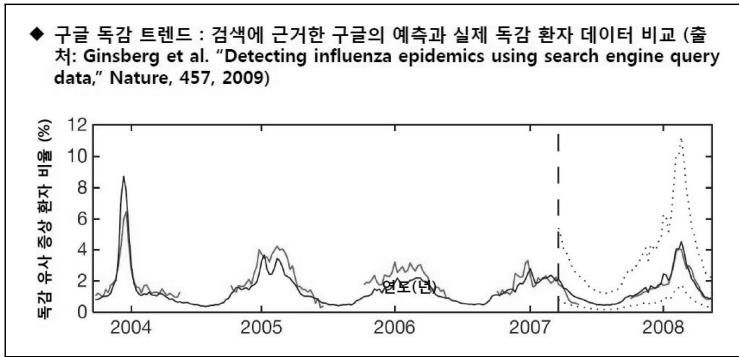
OK 큐피드 사례에서 보여주듯이, 인간의 모호한 삶은 ‘모호한’ 시어로만 이해되는 신비로운 천공이 아니라, 단순히 얽혀 있어 복잡한, 그래서 우리가 풀어나 까발려야 하는 ‘실타래’일 뿐이다. ‘양으로 승부하여 질로 승화한다’는 공대생의 무모한 재도전은 성공적인 것처럼 보인다. 인공족 아틀라스의 사촌 짝인 구글 신은 여기서 더 나아가, ‘빅데이터 기법’을 활용하여 전 세계의 독감 환자 수를 예측할 수 있(다고 주장하)는 ‘빅데이터 분석’을 예시하였다.⁵⁾

OK 큐피드 사례에서와 같이, 독감 환자 수를 예측하기 위해서는 독감 환자를 식별하는 ‘행동 변수’를 찾아내는 것이 중요하다. 특히 ‘우리의 그림자’가 조종하는 ‘행동 변수’의 모습을 보여주는 것이 중요하다. 구글이 고안해 낸 ‘행동 변수’는 아래와 같다.

“사람들은 독감 증상이 있는 경우, ‘기침’, ‘고열’, ‘해열제’ 등과 같은 독감 증상 관련 용어를 검색 엔진을 통해 검색할 것이다.”

5) 구글의 독감 예방 프로그램에 보다 관심 있는 독자는 정하웅·김동섭·이해웅 저의 《구글 신은 모든 것을 알고 있다》(사이언스북스, 2014)를 참조하기 바란다.

그림 4

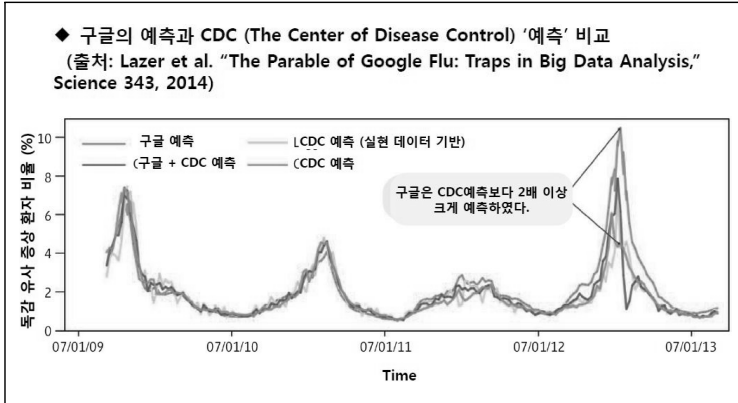


구글 자체가 검색 엔진을 운용하기 때문에, 어느 시점에, 어느 지역에서 독감 증상 관련 검색어가 ‘증폭’하는지 구글 서버를 통해 파악할 수 있고, 또한 데이터로 기록할 수 있다. 실제 구글은 독감 환자 수와 가장 연관성이 높은 검색어 50개를 선정하고, 2003년에서 2007년 사이의 데이터를 수집·분석한 후, 2008년 환자 예측치를 발표하였다. [그림 4]는 구글의 2008년 예측치를 실제 독감 환자 집계 수와 비교한 것이다. [그림 4]에서 보듯이, 구글 예방 프로그램은 거의 완벽하게 독감 환자 수를 예측했다! ‘모호한 천공’을 수많은 숫자로 조합하고 빚어낸, 인공지능 구글로 지탱했을 때, ‘모호한 천공’은 더 이상 그렇게 무거워 보이지 않는다. 구글 신을 숭배하기만 한다면, ‘저만치’ 거리에 있는 20세 여인에게 가슴 설렘, 그러나 그 감정을 시어 ‘저만치’로 세련되게 표현할 재능이 없는 공대생조차 ‘모호한 천공’을 지탱할 수 있음을 보여주는 사례인 셈이다.

이제 구글은 거인족 아틀라스를 대신해, ‘모호한 천공’을 짊어질 수 있어 보인다. 구글은 ‘모호한 천공’ 정도는 한 손가락으로도 지탱할 수 있는 ‘폭력성’을 자랑이라도 하는 듯, 2012년 독감 환자 수 예측치를 발표하였고, 이 예측치는 실제 환자수와 일치할 것이라 장담했다. [그림 5]는 2012년 구글의 예측치와

미국질병통제예방센터가 실제 보고한 독감 환자 수(CDC ‘예측’이라 명명된 것)를 비교한 것이다.

그림 5



구글 신의 장담과는 달리, 구글 독감 환자 수 예측 프로그램은 ‘정확히’ 실제 환자수보다 2배이상을 과대 예측하였다! 삶의 ‘모호함’이라는 무게에 짓눌려 압사해 버린, 피에 버무린 살점이 꽃잎처럼 산화한 사내의 이름은 바로 ‘구글 신’이었다. ‘모호한 천공’은 여전히 버겁고, 그 버거움을 짊어져야 하는 사내는 거인족 아틀라스이지, 인공족 ‘구글 신’이 아님을 보여준다. 그렇다면 이러한 구글 예측 실패는 구글이 고안한 ‘빅데이터 분석’의 실패인가? 아니면 일반적 의미의 ‘빅데이터 분석’의 한계인가? 다음 장에서 이 질문에 대한 답을 찾고자 한다.

‘모형적(模型的) 사고’로서의 ‘빅데이터 분석’

보편적 과학 분석과 달리, ‘우리의 그림자’가 ‘빅데이터 분석’ 과정에 투영되어야 하는 이유는, 관찰의 대상이 분석의 주체인 ‘보편적’ 인간, 바로 우리 자신이기 때문이다. 분석의 주체인 ‘보편적’ 인간이 바로 ‘보편적’ 인간을 관찰·분석 대상으로 삼기에, ‘빅데이터 분석’에서는 항상 분석의 대상과 분석의 주체 간의 거리, 즉 관찰자인 ‘보편적’ 인간과 관찰 대상인 ‘보편적’ 인간 사이의 거리는 시어 ‘저만치’ 정도의 모호한 거리이다. 대부분의 ‘빅데이터 분석’에 관한 저작(著作)들을 보면, 관찰자인 ‘보편적’ 인간과 관찰 대상인 ‘보편적’ 인간 사이의 거리에 ‘저만치’만큼의 모호한 거리가 있다는 것을 인식하지 못하거나, 설사 인식한다 하더라도 단순히 ‘우려’하는 수준에 머물고 있다. 관찰자인 ‘보편적’ 인간과 관찰 대상인 ‘보편적’ 인간 사이에 ‘저만치’ 놓여 있는 모호한 지대(地帶)를 확산시키는 중심축은 소위 ‘빅데이터 분석가’의 ‘모형적(模型的) 사고’에 대한 진지한 고찰의 결여에 있다고 진단할 수 있다. 앞의 절에서 본 구글 독감 예측 프로그램의 과대 예측은 바로 구글의 ‘빅데이터 분석가’가 세심하게 ‘모형적 사고’를 고찰하지 못한 점에 있다고 말할 수 있다.

‘모형적 사고’를 이해하기 위해서는 먼저 ‘모형’이 무엇인지 이해할 필요가 있다. ‘모형’이란, 세계의 관측자로서의 ‘문제의식’을 갖고 있는 관찰자가, 세계의 속성 중 필수 불가결한 요소만 취사선택하는 ‘추상화’의 과정을 통하여 ‘실존하는’ 세계를 ‘수리적 언어’로 축소화, 재구성한 작은 세계(microcosm)를 의미한다. ‘모형적 사고’란 이 축소화되고 재구성된 작은 세계를 가지고 실제 세계를 분석하는 것을 의미한다. 또한 ‘만일 내가 이 축소화된 세계에 살고 있다면 나는 어떠한 행동을 할까? 또는 나는 어떠한 행운과 불행을 맞이할 수 있을까?’ 머릿속으로 상상하는 것을 ‘사고 실험(Das Gedankenexperiment)’이라고 하는데, ‘모형적 사고’는 이 ‘사고 실험’의 활동도 내포한다.

이 ‘모형적 사고’의 근저에는 세계의 관측자로서의 인간이 이 ‘복잡하고 모호한’ 세계를 ‘즉시적’으로 이해하는 것은 불가하지만, 순차적·점진적으로 이해하는 것은 가능할 것이라는 믿음을 암묵적으로 전제하고 있다. 20세기를 대표하는 독일의 수학자 힐버트(David Hilbert)의 묘비명에 적혀 있는 “우리는 알아야 한다. 그리고 우리는 알 것이다.(Wir müssen wissen, Wir werden wissen.)”라는 어구는 바로 ‘모형적 사고’의 기본적, 암묵적 전제를 함축적으로 표현한 것이라 볼 수 있다. 따라서 ‘모형적 사고’란 우리의 삶이 ‘회색’인 것은 알지만, ‘복잡하고 모호한’ 색깔이기에 상대적으로 ‘분명한’ 색깔인 흑색과 백색으로 ‘일단’ 우리의 삶을 이해하는 것을 의미한다. 모호한 회색의 삶을 이해하기 위해, 필수 불가결한 요소인 흑색을 취사선택하는데, 이렇게 흑색을 취사선택하는 과정을 ‘변수화(變數化)’라고 부른다. 즉 취사선택된 필수 불가결한 요소는 변수(變數, variable)로 표현하는 것이다. 반면에 선택받지 못한, 버려지는 삶의 나머지 요소는 상수화(常數化) 처리를 하는데, 상수(常數, constants)라는 말 자체가 의미하듯이, 취사선택되지 못한 나머지 요소는 관찰자의 관찰 유무와는 독립적으로 항상 변하지 않는 것, 일정한 속성을 유지하는 것으로 간주한다. 변수는 크게 ‘행동 변수’와 ‘매개 변수(媒介變數, parameter)’로 나누어질 수 있는데, 행동 변수란 관찰자의 관찰 기간 동안, 계속 관찰되어야 하는 대상이고, 매개 변수란 관찰자의 명백한 관찰 대상이지만, 관찰 기간 동안 일정한 속성을 유지하는 것, 즉 관찰 기간 동안은 변하지 않는 ‘안정성’을 갖는 변수로 이해된다. 모든 과학적 모형은 행동 변수, 매개 변수, 상수 세 가지로 이루어져 있다. ‘모형적 사고’란 결국 행동 변수, 매개 변수, 상수, 이 세 가지 성분으로 세상을 이해하는 것이다. ‘빅데이터 분석’은 우리의 삶을 대상으로 삼기 때문에, ‘빅데이터 분석’에서의 ‘모형’은 한 가지 성분을 더 추가하게 되는데, 앞에서 논의했다시피, ‘우리의 그림자’인 가공의 ‘인격체’가 부여되어야 한다. 이 가공의 ‘인격체’는 바로 모형이 펼쳐내는 축소화된 세계에서, 행동 변수와 매개 변수를 조종하는 우리의 ‘아바타’라 볼 수 있다.

이 인격체인 ‘아바타’의 존재가 순수 과학에서의 모형과 인간의 삶을 다루는, ‘빅데이터 분석’을 포함한 사회 과학에서의 모형과의 차이를 만들어 낸다. 요약하면 ‘빅데이터 분석’에서의 ‘모형적 사고’란 상수, 매개 변수, 행동 변수, 그리고 변수들을 조종하는 가공의 ‘인격체’ 네 가지 성분으로 세상을 이해하는 것이다. 여기서 반드시 지적되어야 할 것은, ‘빅데이터 분석’에서 다루는 모형에서 우리의 ‘아바타’, 즉 가공의 ‘인격체’는 사전적(事前的, ex ante)으로, 또는 선험적(先驗的, a priori)으로 존재하고 행동 변수를 결정하지만, 사후적(事後的, ex post)으로, 또는 경험적(經驗的, a posteriori)으로는 실존하지 않는다는 점이다. 다시 말하면, 우리가 데이터에서 관찰하는 것은 행동 변수이지, ‘인격체’가 아니다. 우리의 ‘아바타’는 사후적으로 관찰된 행동 변수를 통해 귀납적으로 추리(推理)한다. 동시에 연역적으로 행동 변수를 조종하는 우리 ‘인격체’의 의지, 목적, 동기 등을 추론해 ‘인격체’의 행동 강령 ‘제1원칙(the first principle)’을 수립한 후, ‘선험적’으로 우리 ‘인격체’의 존재를 (논리적으로) 증험한다. 귀납적으로 추리하고, 동시에 연역적으로 논증할 때, 우리의 ‘아바타’, 우리의 ‘그림자’, 우리의 ‘일그러지고, 이름 없는 피아니스트’의 실체를 어느 정도 가늠할 수 있을 뿐이다. 그러나 이렇게 파악한 ‘인격체’의 실존조차도 철저히 ‘맥락적(contextual)’이다. 다시 말하면, 가공의 ‘인격체’는 세상 이치의 맥락에 따라, 다양한 몸짓을 보여 줄 뿐이지, 전체의 모습을 절대 보여 주지 않는다. 이 가공의 ‘인격체’가 맥락적일 수밖에 없는 이유는 바로 세상의 회색을 재현하는 것이 아니라, 세상의 흑색만을 재현하는 ‘모형적 사고’ 안에서 또는 모형의 ‘사고 실험’에서만 존재할 수밖에 없기 때문이다. 모형의 용어로 좀 더 ‘정확히’ 표현한다면, 각 모형은 일정하게 변하지 않는 상수항(常數項)을 가지고 있고 또한 관찰 기간 동안은 변하지 않는 ‘안정성’을 갖는 매개 변수를 포함하고 있다. 이러한 불변적인 성질의 ‘안정성’은 우리가 관심 있는 필수 불가결하다고 믿는 요소(행동 변수)를 어떻게 잡을 것인가, 다시 말하면 필수 불가결한 요소를 흑색으로 볼 것인가, 또는 백색으로 볼 것인가, 또한 관찰

시간을 어떻게 잡을 것인가에 따라 정해지기 때문에, 모형 세계는 맥락적일 수밖에 없고, 그 모형 세계에 살고 있는 우리의 ‘아바타’ 또한 맥락적인 존재이다.

서두에서 지적했듯이, ‘빅데이터 분석’에서 부지불식간에 드러나는 관찰자인 ‘보편적’ 인간과 관찰 대상인 ‘보편적’ 인간 사이의 ‘모호한’ 거리는 바로 모형 세계에 사전적으로, 또는 선형적으로 존재할 수밖에 없는 ‘아바타’의 불완전한 실존성, 그리고 맥락적인 관념성에 기인하는 것이라 하겠다. 그리고 이 ‘아바타’의 불완전하고 맥락적인 몸짓은 오로지 ‘모형적 사고’의 진지한 고찰에서만 파악될 수 있다 하겠다.

이제 다시 구글 독감 예측 프로그램 사례로 돌아가자. 구글이 고안한 ‘행동 변수’는 “사람들은 독감 증상이 있는 경우, ‘기침’, ‘고열’, ‘해열제’ 등과 같은 독감 증상 관련 용어를 검색 엔진을 통해 검색할 것이다.”라는 것이었다. 그런 ‘모형적 사고’의 관점에서 보면, 이 행동 변수는 좀 더 세심하게 다루어질 수 있다. 다시 말하면, ‘모형적 사고’를 반영하여 이 ‘행동 변수’를 세심하게 재구성하면 아래와 같다.

“우리의 ‘아바타’는 독감 증상이 있는 경우, ‘기침’, ‘고열’, ‘해열제’ 등과 같은 독감 증상 관련 용어 50개를 검색 엔진을 통해 ‘즉시’ 검색할 것이다.”

또한 이 행동 변수와 더불어 구글은 ‘암묵적’으로 아래의 ‘매개 변수’ 또는 ‘상수항’을 고안했다.

“우리의 ‘아바타’는 독감 증상이 없다면, 독감 증상 관련 용어 50개를 결코 검색하지 않을 것이다.”

다시 말하면, 구글의 예측 프로그램은 적어도 관찰 기간 동안 “우리의 ‘아바타’는 독감 증상이 없다면, 독감 증상 관련 용어 50개를 결코 검색하지 않을 것이다.”라는 ‘아바타’의 행동 강령이 준수된다는 ‘매개 변수’를 부지불식

간에 고안했다. 실제 2008년에 이 ‘매개 변수’는 매우 ‘안정적’이었다. 다시 말하면, 우리의 ‘아바타’가 그러했듯이 실제 독감 증상이 있는 사람들만 검색을 했기에, 구글의 예측 프로그램은 정확했다. 그러나 2012년은 전혀 다른 양상이 펼쳐진다. “우리의 ‘아바타’는 독감 증상이 없다면, 독감 증상 관련 용어 50개를 결코 검색하지 않을 것이다.”라는 이 ‘매개 변수’가 전혀 ‘안정적’이지 않았다. 다시 말하면, 2012년에는 독감 증상이 없는 사람들도 독감 관련 단어를 검색했다! 그렇다면 2012년에는 왜 독감 증상이 없는 사람도 독감 관련 단어를 검색했을까? 그것은 바로 뉴스, 신문과 같은 언론 매체가 구글의 독감 예측 프로그램에서 내놓은 예측치를 보도하기 시작했고, 이 예측치가 보도됨에 따라, 독감에 걸리지 않은 사람들도 다가올 독감을 미리 예방하기 위해 독감 관련 검색어를 검색하기 시작했기 때문이다. 모형으로서의 구글 독감 예방 프로그램은 바로 “우리의 ‘아바타’는 독감 증상이 없다면, 독감 증상 관련 용어 50개를 결코 검색하지 않을 것이다.”라는 매개 변수를 잘못 다루었다고 말할 수 있다. 결론적으로 구글 독감 예측 프로그램은 명백히 ‘모형’이다. 그러나 이 독감 예측 프로그램의 설계자는 이 프로그램이 ‘모형적 세계’를 재현하는, 세상의 축소판이라는 것을 간과한 듯하다. 또한 ‘모형’은 ‘맥락적’이라는 것도 간과한 듯하다. 2008년의 매개 변수와 2012년의 매개 변수가 다를 수 있다는 것을, 그래서 ‘모형’은 맥락적이라는 것을 간과한 것이다. 마지막으로 구글 예측 프로그램은 “프로그램 설계자(관찰자) = 모형 세계의 ‘아바타’ = 실존하는 사람들”이라는 중대한 과오를 범한 듯이 보인다. 이 절의 서두에서 밝혔듯이 관찰자인 ‘보편적’ 인간과 관찰 대상인 ‘보편적’ 인간 사이에는 시어 ‘저만치’ 정도의 모호한 거리가 태생적으로 존재한다. 그리고 그 모호한 거리는 바로 ‘모형적 사고’의 이면일 뿐이다.

결어

한강의 연작 소설 《채식주의자》는 한 편의 음악과도 같은 소설이다. 주인공이 ‘비폭력성’의 상징인 ‘나무’로 변용되어 가건만, 놀랍게도 그 ‘비폭력성’을 성취하도록 돕는 것은 다면 형태의 ‘폭력성’이다. ‘폭력성’은 우리의 살점을 먹어 대는, 천진한 어린 악마가 우리 귀에 속삭이는 방식으로, 또는 바로크 음악의 푸가처럼 ‘폭력’이라는 일관성 있는 주제를, 여러 성부에 있는 화자의 입을 통해 노래하는 방식으로 전달된다. ‘폭력성’은 때로는 무지로, 때로는 무관심으로, 때로는 ‘꽃’과의 ‘모호한’ 거리 ‘저만치’를 ‘명백하게’ 파괴하는 피에 버무린 살점을 꽃잎처럼 산화하는 사내의 모습으로 등장한다. 주인공 또한 ‘폭력성’의 주제를 또 다른 성부에서, 자기 학대, 소통 부재의 ‘폭력성’의 마법에 걸린 것처럼 노래한다. 그러나 이 다양한 성부에 기거하는 ‘폭력성’은 주인공을 ‘십자가’ 형태의 나무가 있는 골고다 언덕으로, 마치 예정된 길을 인도하듯이 인도한다. 그럼에도 ‘진흙 속의 연꽃’이라는 ‘상투적인’ 노래, 즉 ‘명백히’ 한 가지 으뜸음을 가진, 조성(調聲, tonality)을 가진 음악을 허락하지 않는다. 오히려 1721년에 출간된 바흐(Johann Sebastian Bach)의 《평균율》 1권 중 마지막 푸가에서, 인류 역사상 최초의 12음 기법으로 작성된, 무조(無調, atonality) 음악의 가능성을 목도한 쇤베르크(Arnold Schönberg)의 환희를, 《채식주의자》는 재현한다. 《채식주의자》는 독자로 하여금 으뜸음을 허용하지 않는, 무조성(無調聲, atonality)의 소리를 발견하는 ‘쇤베르크의 그 환희’를 만끽하게 해 준다.

‘빅데이터 분석’ 또한 우리의 삶을 한 편의 음악처럼 들려준다. ‘빅데이터’가 들려주는 음악은 그냥 지나치기는 아쉬운 거리, 그러나 내 손길을 뻗친다 한들, 인연의 끈을 잡을 수 없는 거리, 아련히 다가오는, 그러나 이내 시나브로 멀어져가는 ‘복학생 오빠’와의 거리, 그래서 가슴 설레는, 한(恨)이 이슬처럼 망울지는, ‘마음’의 ‘모호함’을 드러내는 ‘저만치’를 으뜸음으로 하는 음악이

아니다. 소설 《채식주의자》처럼, ‘빅데이터’를 움직이는 동인(動因)은 다양한 형태의 ‘행동 변수’가 보여 주는 ‘폭력성’이다. ‘명확히’ 대칭적이고 균등한 가치를 가진 12개의 ‘폭력적인’ 음은 ‘저만치’라는 으뜸음을 허락하지 않는다. 12개의 음은 으뜸음 ‘저만치’를 분석(分析)의 이름으로 갈기갈기 찢어버리고, 까발린다. 바로 쇤베르크의 12음 기법에 의한 무조 음악의 이상을 삶이라는 악보에 기록하는 것, 이것이 ‘빅데이터’의 기본 성향이다. 20세기 최고의 바흐 건반 음악 연주자인 글렌 굴드(Glenn Gould)의 연주에서도 이러한 ‘빅데이터’의 ‘시체 해부자’ 성향을 볼 수 있다. 글렌 굴드의 주법은 고적한 산사의 묵탁 소리와도 같은 ‘명확한’ 언어를 전달한다. 굴드는 특히 그냥 지나치기는 아쉬운 거리, 그러나 내 손길을 뻗친다 한들, 인연의 끈을 잡을 수 없는 거리, 아련히 다가오는, 그러나 이내 시나브로 멀어져가는 ‘복학생 오빠’와의 거리, 그래서 가슴 설레는, 한(恨)이 이슬처럼 망울지는, ‘마음’의 ‘모호함’을 드러내는 ‘저만치’를 으뜸음으로 하는 낭만적인 피아노곡의 얼굴을 묵탁 소리와 같은 점묘법을 통해 가차 없이 지워 버린다. 곡의 ‘민낯’을 여과 없이 까발린다. 남겨지는 것은 회색빛 불교의 승려복 그리고 머리카락 하나 없는 두상뿐이다. 그림에도 굴드는 멈추지 않는다. 굴드는 ‘명확한’ 묵탁 소리로 언제나 해체하고 싶은 시체 해부자의 본능에 시달렸다. 실제 ‘바흐’ 연주자로 유명하지만, 12음 기법으로 작성된 쇤베르크의 곡 또한 너무나 잘 연주하였다. 그러나 끊임없는 해부의 본능으로 탄생한 쇤베르크의 곡을 듣고 있노라면, ‘모호함’을 제거한 후에 삼킨 ‘명확한’ 환희가 아니라, 해체하고 나니, 아무것도 없다는 ‘공허함’만이 몰려온다. 오직 시체 해부자의 본능에 시달리는 ‘빅데이터 분석’은 ‘모형적 사고’를 진지하게 고찰해 본 적이 없는 서툰 ‘빅데이터 분석가’의 모습이지, ‘빅데이터’의 진면모와는 거리가 멀다. ‘분명히’ ‘모형적 사고’는 또 다른 차원의 ‘저만치’라는 ‘모호한’ 거리를 만들어 낸다. 《채식주의자》는 폭력적이지만, 삶의 색채를 지워 버리지 않는다. 진짜 ‘빅데이터 분석’은 폭력적이지만, 우리 삶의 색채를 더욱 다채롭게 만들어 줄 것이다. 삶에서 무조성의 소리를

발견하는 쇠베르크의 환희는 우리의 모호한 삶을 더욱 모호하면서도 분명하게 이해하게 해 줄 것이다.

마지막으로 ‘모형적 사고’를 가장 잘 이해했던 사람은 시인 무리에 있다. 하나의 몸짓 외에는 명확히 서술할 수 없는 ‘모호한’ 삶에 ‘명확’하고 ‘폭력적인’ 언어의 이름을 붙임으로써 ‘저만치’ 거리에 있는 존재를 ‘꽃’으로 확정해 버리는 것, 바로 회색에서 필수 불가결하다고 믿는 것에 흑색이라는 명확한 모형을 설정해 버리는 행위를 시인은 아래와 같이 읊조린다.⁶⁾

내가 그의 이름을 불러주기 전에는
그는 다만
하나의 몸짓에 지나지 않았다.
내가 그의 이름을 불러주었을 때
그는 나에게로 와서
꽃이 되었다.

6) 김춘수의 시 <꽃> 중에서.

