Sign Languages and Linguistics Convergence Research
Vol. 2. No. 2 (2025) pp.029~061

멀티모달 말뭉치의 내적 · 외적 타당성 고찰

조 용 준 · 안 희 돈*

(건국대학교)

Cho, Yongjoon & Ahn, Hee-Don. 2025. Internal and External Validity of Multimodal Corpora: A Critical Examination. ***Sign Languages and Linguistics Convergence Research,*** *2*(2), 29-61. This study aims to critically reinterpret the history of multimodal corpora through the lens of the epistemological tension between two key paradigms for constructing scientific knowledge: 'control' for pursuing internal validity and 'naturalness' for pursuing external validity. To this end, this paper first systematically outlines the philosophical foundations of science and methodological criteria for each paradigm. Based on this theoretical framework, it analyzes cases of the control paradigm, represented by M2VTS and IEMOCAP, and cases of the naturalness paradigm, such as FOLK, CEJC, and CANDOR. It also examines the characteristics of contemporary corpora like MISP, where the two paradigms intricately intersect. Furthermore, this analysis is extended beyond the audio-vocal channel to the field of sign language corpora, which centers on the visual-motor channel, to analyze the unique methodological dilemmas faced in sign language corpus linguistics. Finally, synthesizing the entire discussion, this study proposes an analytical framework presenting three potential design paths (maximization of internal validity, maximization of external validity, or a compromise) for future corpus researchers to consider according to their research objectives.

Key Words: Multimodal Corpora, Naturalness, Controlled Data, Internal Validity, External Validity, Ecological Validity, Sign Language Corpora

* 주저자: 조용준, 교신저자: 안희돈

1. 서론

인간의 의사소통은 본질적으로 멀티모달적이다. 우리는 단어뿐만 아니라 제스처, 시선, 표정, 억양 등 여러 모달리티를 동시에 활용하여 의미를 구성하고 관계를 맺는다. 이러한 관점에 기반하여, 지난 수십 년간 언어학 및 인공지능 연구 분야에서는 인간 상호작용의 총체적 모습을 포착하기 위해 수많은 멀티모달 말뭉치가 구축되어 왔다. 그러나 이러한 양적 성장에도 불구하고, 말뭉치를 구축하는 방법론의 기저에 놓인 철학적 기반에 대한 논의는 상대적으로 부족했다. 멀티모달 말뭉치의 발전사는 단순히 기술의 진보를 반영하는 것을 넘어, "무엇이 좋은 데이터인가?"라는 질문에 답하기 위한 서로 다른 과학적 이상(理想) 사이의 지속적인 긴장 관계를 드러낸다.

본 연구의 핵심적인 문제의식은 이 긴장 관계가 연구의 신뢰도를 담보하는 두 가지 상충하는 기준, 즉 내적 타당성(internal validity)을 추구하는 '통제된 데이터(controlled data)' 패러다임과 외적 타당성(external validity)을 지향하는 '자연성(naturalness)' 패러다임의 대립에서 비롯된다는 점에 있다. 본고의 목적은 멀티모달 말뭉치의 발전사를 이 두 패러다임 사이의 인식론적 긴장 관계라는 렌즈를 통해 비판적으로 재해석하는 것이다. 이를 통해, 향후 멀티모달 말뭉치를 설계하려는 연구자들이 자신의 연구 목적에 맞춰 어떤 방법론적 선택을 할 수 있는지, 그리고 각 선택이 어떤 의미와 한계를 갖는지에 대한 분석적 프레임워크를 제시하고자 한다.

이를 위해, 본고는 다음과 같은 순서로 논의를 전개한다. 2장에서는 본론의 분석 대상이 될 주요 멀티모달 말뭉치들의 역사적 흐름을 세대별로 간략히 개관한다. 3장에서는 '통제'와 '자연성'이라는 두 패러다임의 과학철학적 기반과 구체적인 방법론적 기준들을 심층적으로 고찰한다. 4장에서는 3장에서 제시한 이론적 틀을 사용하여 주요 음성언어 말뭉치들을 비판적으로 재평가하며, 5장에서는 이 분석을 시각 언어라는 특수한 과제를 안고 있는 수어 말뭉치에까지 확장한다. 마지막으로 6장에서는 전체 논의를 종합하여, 연구 목적에 따른 말뭉치 설계의 가능성을 프레임워크 형태로 제시하며 결론을 맺을 것이다.

2. 멀티모달 말뭉치의 시대별 흐름

멀티모달 말뭉치의 발전사는 기술의 진보와 연구 패러다임의 전환을 명확하

게 보여준다. 1990년대 후반 고도로 통제된 환경에서 융합의 가능성을 탐색하던 초기 단계를 지나, 2000년대에는 복잡한 인간 상호작용과 감성 분석으로 그 영역이 확장되었다. 2010년대에 들어서는 '현실 세계(in the wild)'에서의 자연스러운 데이터를 포착하려는 움직임이 주류를 이루었고, 2020년대 현재는 전례 없는 규모의 데이터에 공학적 난제 해결이라는 뚜렷한 목적성을 결합하는 방향으로 나아가고 있다. 본 장에서는 이러한 흐름을 대표하는 주요 말뭉치들을 세대별로 간략히 개관하여, 4장에서 이루어질 심층 분석의 토대를 마련하고자 한다.

2.1. 제1세대: 통제된 환경에서의 융합 가능성 탐색

멀티모달 말뭉치 연구의 태동기는 단일모달(unimodal) 기술이 가진 한계를 극복하기 위한 대안으로서 정보 융합의 잠재력을 탐색하는 데 집중되었다. 이 시기를 대표하는 말뭉치는 M2VTS(Multi Modal Verification for Teleservices and Security Applications)와 그 확장판인 XM2VTSDB이다(Pigeon & Vandendorpe, 1997; Messer et al., 1999). 벨기에 루뱅 가톨릭 대학교(UCL)에서 수집된 이들 말뭉치는 참가자들이 자신의 모국어(대부분 프랑스어)로 숫자를 세는 음성 데이터와 정면 얼굴 및 회전 영상 데이터를 포함한다(Pigeon & Vandendorpe, 1997). 특히 확장판인 XM2VTSDB는 295명의 참가자를 대상으로 구축되었으며, 클라이언트-사칭자 비율과 인구통계학적 변인을 고려한 체계적인 수집 프로토콜을 제시한 최초의 대규모 생체인식 말뭉치라는 특징을 갖는다(Messer et al., 1999).

이 프로젝트들은 "서로 다른 생체인식 모달리티는 독립적인 정보를 제공하며, 이들의 적절한 결합은 개별 모달리티의 약점을 상호 보완할 수 있다"는 핵심적인 가정에 기반했다(Pigeon & Vandendorpe, 1997). 이러한 가설을 순수하게 검증하기 위해, 데이터 수집 과정은 극도로 통제된 실험실 환경에서 이루어졌다. 연구팀은 배경, 조명, 참가자의 발화 내용('0'부터 '9'까지 세기), 그리고 머리의 움직임까지 철저히 통제된 표준화된 과업을 설계했다. 데이터는 고품질 비디오 카메라로 촬영되었으나, 저장 공간의 한계로 인해 더 낮은 해상도의 형식으로 변환되기도 했다(Pigeon & Vandendorpe, 1997). 이는 자연스러운 인간 행동을 포착하려는 목적이 아니라, 오직 '융합 알고리즘' 자체의 성능을 외적 변수의 개입 없이 평가하기 위한 실험적 통제 장치였다.

M2VTS에서 얻은 경험을 바탕으로 구축된 XM2VTSDB는 단순히 참가자 수를 늘리는 것을 넘어, 방법론적 정교함을 더했다. 특히 개인 인증 시스템의 실용성을 평가하기 위해선 충분한 규모의 '사칭자(imposter)' 데이터가 필수적이라

는 인식하에, 시스템의 일반화 성능을 검증할 수 있도록 데이터베이스를 설계했다(Messer et al., 1999). 이처럼 1세대 말뭉치들은 데이터 자체의 언어학적, 사회과학적 가치보다는, 그 데이터를 통해 무엇을 '가능하게 했는가'에 그 의의가 있다. 즉, 이 말뭉치들은 멀티모달 생체 인식이 공학적으로 실현 가능한 분야임을 최초로 증명하는 실험의 장을 제공했으며, 이는 이후 더 복잡하고 자연스러운 상호작용을 분석하려는 후속 연구들의 단단한 발판이 되었다.

2.2. 제2세대: 복잡한 상호작용 분석으로의 확장

2000년대 중반, 멀티모달 연구의 관심은 통제된 환경에서의 개인 식별을 넘어, 여러 사람이 참여하는 역동적이고 복잡한 상호작용의 본질을 이해하는 것으로 확장되었다. 이러한 패러다임의 전환을 상징하는 기념비적인 결과물이 바로 AMI(Augmented Multi-party Interaction) Meeting Corpus이다(Carletta et al., 2005). 총 100시간 분량의 회의 녹화 데이터로 구성된 이 말뭉치는, 작업 그룹의 효율성을 향상시키는 "회의 브라우징 기술" 개발을 명시적인 목표로 삼았다(Carletta et al., 2005).

AMI 말뭉치의 가장 큰 특징은 데이터 수집 방법론에 있다. 연구팀은 여러 대의 카메라와 근거리/원거리 마이크 어레이, 전자 화이트보드, 디지털 펜 등 다양한 센서가 정밀하게 동기화되는 특수 제작 '계측화된 회의실(instrumented meeting rooms)'을 구축했다(Carletta et al., 2005). 또한 참가자들이 명확한 목표를 가지고 현실적인 상호작용에 몰입하도록 '시나리오 기반 유도'라는 설계 방식을 채택했다. 참가자들은 가상의 전자 회사 직원 역할을 부여받아 새로운 TV 리모컨을 개발하는 과업을 수행했는데, 이는 목적 지향적 상호작용을 유도하면서도 연구자가 회의의 성과를 분석하기 용이한 환경을 제공했다(Carletta et al., 2005).

이처럼 AMI 말뭉치는 순수 공학 분야였던 멀티모달 연구가 화행(dialogue acts)이나 그룹 동역학(group dynamics)과 같은 언어학, 사회학, 심리학의 개념을 적극적으로 통합하는 학제간 과학으로 진화하는 중요한 계기를 마련했다. 연구의 질문이 "이 사람이 허가된 사용자인가?"에서 "이 회의에서 어떤 결정이 내려졌는가?"로 근본적으로 변화했음을 보여주는 사례이며, 이후 복잡한 인간 상호작용을 연구하는 데 있어 중요한 방법론적 혁신으로 평가된다.

2.3. 제3세대: 감성 및 비언어적 표현의 정밀 측정

복잡한 다자간 상호작용 분석으로 연구의 외연이 확장된 이후, 2000년대 후

반 연구자들은 인간 의사소통의 가장 미묘하고 도전적인 영역인 '감성(emotion)'을 깊이 파고드는 새로운 방향으로 나아갔다. 이 흐름을 대표하는 IEMOCAP(Interactive Emotional Dyadic Motion Capture Database)은 음성과 제스처의 통합적 분석을 통해 표현적 인간 의사소통(expressive human communication) 연구를 촉진하는 것을 목표로 했다(Busso et al., 2008). 10명의 전문 배우가 2인(dyadic)으로 짝을 이뤄 상호작용을 수행했으며, 이를 통해 약 12시간 분량의 시청각 데이터를 구축했다(Busso et al., 2008).

IEMOCAP의 가장 핵심적인 기술적 혁신은 고정밀 모션 캡처 기술을 사용하여 미세한 비언어적 신호를 정량화한 데 있다. 연구팀은 VICON 모션 캡처 시스템을 이용해 배우의 얼굴에 53개의 마커를 부착하여, 초당 120 프레임의 속도로 얼굴 표정과 제스처의 3차원 움직임을 정밀하게 기록했다(Busso et al., 2008). 또한 감성 연구에 필요한 특정 발화들을 효과적으로 수집하기 위해, 배우들에게 사전에 엄선된 연극 대본을 연기하게 하는 '대본 기반 세션(Scripted Sessions)'과 특정 감정을 유발하도록 설계된 가상의 시나리오 안에서 자유롭게 즉흥 연기를 펼치게 하는 '즉흥 연기 세션(Spontaneous Sessions)'을 병행했다(Busso et al., 2008).

수집된 데이터는 여러 평가자에 의해 분노, 행복, 슬픔과 같은 범주형 레이블과, 유인가(valence), 각성(activation), 지배성(dominance)과 같은 차원형 레이블로 상세하게 주석되었다(Busso et al., 2008). '감성'이라는 주관적인 현상에 대해 높은 데이터 충실도와 신뢰도 높은 정답(ground-truth) 품질을 확보하려는 이러한 방법론적 선택을 통해, IEMOCAP은 현재까지도 멀티모달 감성 인식 및 감성 컴퓨팅(Affective Computing) 분야의 핵심적인 벤치마크로 널리 사용되고 있다.

2.4. 제4세대: 실제 환경에서의 자발적 상호작용 포착

2010년대에 들어서면서 멀티모달 말뭉치 연구는 중요한 전환점을 맞이했다. 이전 세대의 연구들이 실험실이나 특정 시나리오와 같이 어느 정도 정해진 틀 안에서 데이터를 수집했던 것과 달리, 4세대는 연구의 무대를 '현실 세계(in the wild)'로 옮겨왔다. 이 시기의 핵심 목표는 실험실의 제약을 벗어나 실제 일상 환경에서 발생하는 예측 불가능하고 자발적인 상호작용을 대규모로 포착하여, 실제 언어 사용의 다양성을 포괄적으로 기술하는 것이었다. 독일어권의 FOLK (Forschungs- und Lehrkorpus Gesprochenes Deutsch)와 일본어권의 CEJC(Corpus of Everyday Japanese Conversation)가 이러한 흐름을 대표한다(Schmidt, 2014;

Koiso et al., 2018).

독일의 FOLK 말뭉치는 독일어 구어의 모든 변이형을 체계적으로 기록하는 대규모 참조 코퍼스 구축을 목표로 했다(Schmidt, 2014). 이 목표를 달성하기 위해, 연구팀은 데이터의 다양성을 확보하는 데 주력했다. 사적인 영역(가족 대화, 친구 모임), 제도적 영역(학교 수업, 직장 회의), 그리고 공적인 영역(서비스 상황, 의료 상담) 등 다양한 상황적 맥락을 포괄했으며, 독일 내 여러 지역 방언과 연령, 교육 수준, 직업 등 다양한 사회적 배경을 가진 화자들을 균형 있게 포함하고자 했다(Schmidt, 2016). 이는 특정 과업 수행이 아닌, 언어 공동체의 실제 언어생활을 총체적으로 담아내려는 시도였다.

일본의 CEJC 말뭉치역시 실제 사회적 행동의 다양성을 포착하는 균형 잡힌 대규모 일상대화 말뭉치 구축을 목표로 했다(Koiso et al., 2018; Koiso et al., 2022). CEJC의 가장 큰 특징은 데이터 수집의 어려움을 극복하기 위해 도입한 혁신적인 이중 접근법에 있다. 첫째는 '개인 기반 수집(Individual-based recording)'으로, 모집된 일반인 정보 제공자가 휴대용 녹화 장비를 소지하고 수개월에 걸쳐 자신의 가정이나 직장 등 실제 생활 공간에서 대화를 직접 녹화하는 방식이다. 둘째는 '특정 장면 수집(Specific-scene recording)'으로, 개인 기반 방식으로는 수집이 어려운 특정 상황의 데이터를 연구팀이 직접 녹화하여 보완했다(Koiso et al., 2018). 이처럼 4세대 말뭉치들은 실제 언어 사용의 모습을 최대한 왜곡 없이 담아내기 위한 방법론적 고민을 통해, 이후 언어 연구의 방향을 크게 바꾸어 놓았다.

2.5. 제5세대: 대규모화와 복합적 문제 해결

2020년대에 들어 멀티모달 말뭉치는 전례 없는 규모와 뚜렷한 목적성을 결합하는 방향으로 발전하고 있다. 이 시기의 말뭉치들은 단순히 데이터를 수집하여 제공하는 것을 넘어, 심리학적 상호작용의 근본 원리를 탐구하거나, 현실 세계의 복잡한 공학적 난제를 해결하기 위한 대규모 '실험장(Testbed)'으로서의 성격을 띤다. CANDOR(Conversation: A Naturalistic Dataset of Online Recordings)와 MISP(Multi-modal Information based Speech Processing) 챌린지 말뭉치는 이러한 최신 흐름의 두 가지 다른 방향을 대표한다.

CANDOR는 인간 상호작용에 대한 학제간 연구를 촉진하기 위해 구축된 거대 말뭉치로, 낯선 사람들 간의 비대면 영상 채팅으로 이루어진 1,650개 이상의 대화, 총 850시간 이상에 달하는 방대한 데이터를 포함한다(Reece et al., 2023).

이 말뭉치의 가장 독창적인 특징은 말차례와 같은 저수준의 행동 패턴을 참여자의 주관적인 심리 상태와 직접적으로 연결했다는 점이다. 이를 위해 연구팀은 대화 데이터와 함께, 참여자들의 성격 특성, 대화 전후의 기분 변화, 그리고 상대방과 대화에 대한 상세한 평가 등 광범위한 설문 데이터를 수집했다(Reece et al., 2023). 이는 미시적 상호작용 패턴이 인간의 심리와 관계 형성에 미치는 영향을 대규모 데이터로 분석할 수 있는 기반을 마련했다.

이와는 다른 축에서, MISP챌린지 말뭉치는 어렵고 도전적인 현실 세계 시나리오에서 음성 처리 기술의 성능을 한계까지 밀어붙이고 객관적인 벤치마크를 제공하기 위해 명시적으로 설계되었다(Wang et al., 2023). 이 프로젝트는 2명에서 6명의 사람들이 TV 소음이 상당한 배경에서 자유롭게 대화하는 중국어 '가정 내 TV 시청 시나리오'를 데이터로 수집했다. 이는 원거리 마이크로 인한 반향, 여러 사람의 발화 겹침, 저품질 비디오와 같은 복합적인 문제 상황을 의도적으로 포함하며, 현재 기술로 해결하기 어려운 문제 상황을 제공한다(Wang et al., 2023). MISP는 '화자 분할 및 인식'이라는 명확한 공학적 문제 해결을 목표로 한다는 점에서, 특정 목적을 위해 현실의 복잡성을 그대로 활용하는 현대 말뭉치의 새로운 경향을 잘 보여준다.

3. 통제된 데이터와 자연성의 문제

3.1. 자연성과 생태학적 타당성

코퍼스 언어학의 역사는 언어학자의 내성적 직관에 대한 대안으로서 실제 발화 데이터의 가치를 주장해온 과정이라고 해도 과언이 아니다. 특히 사회언어학이나 담화분석과 같은 분야에서, 통제되지 않은 현실 세계에서 나타나는 '자연 데이터(natural data)'의 중요성은 핵심적인 대전제로 받아들여진다.

이러한 연구의 궁극적인 목표는 연구 결과가 실험실을 넘어, 실제 생활 세계, 즉 인간 행동의 '생태계(ecology)'에 얼마나 일반화될 수 있는지를 확보하는 데 있다. 이 일반화 가능성의 정도를 '생태학적 타당성(ecological validity)'이라고 한다. 이 개념은 '연구 결과가 현실을 얼마나 잘 설명하고 예측할 수 있는가?'라는 질문과 직결되며, 주로 심리학이나 인간-컴퓨터 상호작용(HCI) 분야에서 발전했지만 '실제 언어 사용'을 탐구하려는 언어학의 목표와도 정확히 일치한다.

그렇다면 이 '생태학적 타당성'이라는 목표는 어떻게 달성될 수 있는가? 높은 생태학적 타당성을 확보하기 위해서는, 데이터가 두 가지 핵심적이면서도 서로 구별되는 조건을 충족해야 한다.

첫 번째 조건은 '대표성(Representativeness)'이다. 이는 데이터 '표집(sampling)'의 문제이다. 이 관점은 Brunswik(1956)과 Sinclair(1991)에 의해 강조되었으며, '수집된 데이터 집합이 연구하고자 하는 현실 세계(예: 특정 언어 공동체)의 다양한 조건들(예: 장르, 화자, 상황)을 통계적으로 얼마나 잘 반영하는가?'를 묻는다.

두 번째 조건은 '자연성'이다. 이는 개별 데이터 질(quality)의 문제이다. 이 관점은 Labov(1972)의 '관찰자 역설'이나 대화분석의 '자연 발생적 상호작용(naturally occrring interaction)' 원칙에서 극대화된다(Sacks, 1992). 이는 '개별 데이터가 연구자의 개입이나 관찰 효과(카메라의 존재)로 인해 왜곡되지 않았는가?', 그리고 '그 상호작용이 연구와 무관하게 참여자들의 실제 삶 속에서 실제적인 이유로 어차피 일어났을 상호작용인가?'를 묻는다.[1)]

이 두 조건, 즉 자연성과 대표성은 명백히 다르며, 서로 독립적으로 충족되거나 결여될 수 있다. 예를 들어, 한 가족의 저녁 식사 대화 100시간을 몰래 녹음한 데이터는 자연성은 극도로 높지만, 그 사회 전체의 언어 사용을 반영하지는 못하므로 '대표성'은 매우 낮다. 반대로, 다양한 연령과 직업의 사람들을 인구 비율에 맞게 층화적으로 표집하여 스튜디오에서 특정 대본을 읽게 한 데이터는 대표성은 높을 수 있으나 자연성은 매우 낮다.

이 두 조건, 즉 대표성과 자연성은 '자연성' 패러다임, 즉 외적/생태학적 타당성을 추구하는 연구의 주요 축을 이룬다. 이는 3.2절에서 다룰, 명확한 인과관

1) 본고에서는 자연성과 진정성(authenticity)을 구분한다. 진정성은 그 상호작용이 연구자나 과업 때문에가 아니라 참여자들의 일상적 필요 때문에 일어났는지를 보는 개념이다. 대화분석에서 말하는 자연 발생적 상호작용이 여기에 해당한다(Sacks, 1992). 실제 가족 식사나 민원・상담 통화처럼 연구가 없어도 일어났을 상호작용은 진정성이 높다. 자연성은 관찰・녹화・면접 같은 연구 상황 때문에 발화 양식이 평소와 다르게 조정되지 않았는지를 묻는다. Labov가 관찰자 역설에서 지적했듯이, 관찰이 개입하면 화자가 스타일을 올리거나 말하기를 간소화하는 일이 발생한다(Labov, 1972). 이런 조정이 일어날수록 자연성은 떨어진다.
자발성(spontaneity)은 이 자연성의 주요 구성 요소 가운데 하나다. 대본을 읽었는지, 역할 연기를 했는지, 플랫폼이 발화 구조를 미리 정해놓았는지에 따라 발화는 덜 자발적으로 된다. IEMOCAP처럼 감정 연기를 통해 수집한 자료가 자연성에서 낮게 평가되는 이유가 여기에 있다(Busso et al., 2008). 반대로 일상대화나 현장 기록처럼 발화의 시점과 내용이 참여자에게서 즉시 나오는 경우에는 자발성이 높다. 현실의 일상자료는 보통 (1) 실제로 일어났어야 할 상호작용이면서(진정성), (2) 관찰 때문에 말이 크게 달라지지 않았고(자연성), (3) 대본 없이 산출되었기 때문에(자발성) 이 세 기준을 함께 만족하는 경향이 있다. 이 때문에 이후 장에서는 이 셋이 겹치는 구간을 편의상 '자연성'으로 묶어 부르기로 한다. 다만 이렇게 묶어도 대표성과는 별개다(Sinclair, 1991; Brunswik, 1956).

계 규명을 위해 ‘내적 타당성’을 극대화하는 ‘통제된 데이터’ 패러다임과 대립각을 세운다.

본 장(3장)의 목적은 이 두 패러다임의 철학적 기반을 명확히 정립하는 것이다. 3.2절에서는 ‘통제’ 패러다임의 논리를, 그리고 3.3절에서는 이 두 패러다임 간의 근본적인 ‘인식론적 긴장’을 다룰 것이다. 마지막으로 3.4절 이하에서는, 본 절(3.1)에서 정립한 ‘자연성’과 ‘대표성’이라는 두 축을 실제로 확보하기 위한 구체적인 방법론(예: 라보프의 관찰자 역설, 대화분석의 원칙 등)을 심층적으로 고찰할 것이다.

3.2. ‘통제된 데이터’와 내적 타당성

멀티모달 말뭉치 연구의 초기 역사는 언어학보다는 전산학과 심리학의 방법론에 깊이 뿌리내리고 있다. 이들 학문에서 연구의 신뢰도를 확보하는 전통적인 방식은 명확한 가설을 설정하고, 그 가설에 영향을 줄 수 있는 외적 변수들을 최대한 배제한 후 실험을 설계하는 것이었다. 이러한 연역주의적(hypothetico-deductivism) 연구 전통 하에서 구축된 것이 바로 ‘통제된 데이터(controlled data)’기반의 말뭉치들이다.

‘통제된 데이터’란, 연구자가 측정하고자 하는 핵심 요소를 제외한 다른 모든 변수들을 의도적으로 통제하고 배제하여 수집한 데이터를 의미한다. 이는 연구 가설이나 알고리즘의 성능을 ‘순수하게’ 검증하기 위해, 현실 세계의 복잡성을 일종의 ‘노이즈(noise)’로 간주하고 이를 인위적으로 제거한 결과물이다. 이러한 접근은 실험 결과의 ‘내적 타당성(internal validity)’, 즉 측정된 결과가 다른 외적 요인이 아닌 순전히 실험적 처치에 의한 것임을 보증하는 것을 최우선 목표로 삼는다.

이러한 패러다임을 명확히 보여주는 대표적인 사례가 초기 생체 인식 연구를 위해 구축된 M2VTS 말뭉치다(Pigeon & Vandendorpe, 1997). 이 프로젝트의 목표는 “음성과 얼굴 정보를 융합하면 개인 식별의 정확도가 향상될 것”이라는 명확한 공학적 가설을 검증하는 것이었다. 이를 위해 연구자들은 배경, 조명, 참가자의 발화 내용(‘0’부터 ‘9’까지 세기), 심지어 머리의 움직임까지 철저히 통제된 환경을 구축했다. 현실에서 나타날 수 있는 다양한 변수들을 제거함으로써, 오직 ‘멀티모달 융합 알고리즘’ 자체의 성능을 순수하게 평가할 수 있는 데이터를 확보하고자 한 것이다.

감성 컴퓨팅 분야의 발전에 결정적 기여를 한 IEMOCAP 말뭉치 역시 ‘통제

된 데이터' 패러다임의 연장선에 있다(Busso et al., 2008). 연구팀은 "특정 감정은 특정 비언어적 신호와 체계적으로 연결된다"는 가설을 탐구하기 위해, 전문 배우들에게 특정 감정을 유발하는 대본을 연기하도록 했다. 이는 '감정'이라는 매우 주관적이고 복잡한 현상을 객관적으로 측정 가능한 데이터로 변환하기 위해, 완전한 자연성을 일부 희생하는 대신 데이터의 충실도와 정답(ground-truth)의 품질을 확보하려는 의도적인 선택이었다.

결론적으로 '통제된 데이터' 패러다임은 명확한 가설 검증을 목표로 현실의 복잡성을 의도적으로 단순화한, 전통적 과학 실험의 철학을 반영하는 연구 방법론이라 할 수 있다.

과학적 탐구의 역사에서 가장 강력한 패러다임 중 하나는 특정 현상의 원인을 규명하려는 열망, 즉 인과관계(causality)의 규명에 있다. 이러한 목표를 달성하기 위해 발전된 것이 바로 실험설계(experimental design) 방법론이며, 여기서 수집되는 '통제된 데이터'는 연역주의적(hypothetico-deductive) 과학철학의 산물이다. 연구자는 "X가 Y를 유발한다"는 가설을 설정한 뒤, Y에 영향을 미칠 수 있는 무수한 잠재적 변수들 중에서 오직 X의 영향력만을 순수하게 관찰하고자 한다. 이를 위해 다른 모든 외생 변수(exogenous variable)의 효과를 의도적으로 억제하거나 일정하게 유지하는데, 이 과정을 통해 확보된 데이터가 바로 '통제된 데이터'이다.

이 패러다임에서 데이터의 품질을 결정하는 가장 중요한 척도는 '내적 타당성(internal validity)'이다. 내적 타당성이란, 관찰된 결과가 제3의 알려지지 않은 변수가 아닌 오직 연구자가 가한 실험적 처치(treatment)에 의해서만 비롯되었음을 확신할 수 있는 정도를 의미한다. 학계에서는 이러한 내적 타당성을 확보하기 위해 정교한 기준들을 발전시켜왔다. 대표적으로, 연구 대상을 실험 집단과 통제 집단에 무작위로 배정하는 무작위 할당(randomization)은 관찰되지 않은 개인적 특성이나 외부 요인들이 두 집단에 체계적으로 영향을 미칠 가능성을 통계적으로 최소화하는 핵심적인 절차이다. 또한 모든 참가자가 동일한 환경과 절차를 경험하도록 하는 연구 절차의 표준화(standardization)나, 연구자와 참가자의 기대 심리가 결과에 미치는 영향을 차단하는 눈가림(blinding)기법 역시 내적 타당성을 위협하는 잠재적 '노이즈'를 제거하기 위한 필수적인 장치로 간주된다.

결국 '통제된 데이터'는 현실 세계의 복잡성을 의도적으로 단순화하고 정화하여, 명확한 인과적 법칙을 발견하려는 과학적 열망을 반영한다. 이는 특정 조건 하에서 재현 가능한(replicable) 지식을 생산하는 데 매우 효과적인 방법론이다.

3.3. 인식론적 긴장: 내적 타당성과 외적 타당성의 대립

‘통제된 데이터’와 ‘생태학적 타당성’을 둘러싼 논의의 핵심에는, 과학적 지식의 정당성을 확보하는 두 가지 상충하는 기준인 ‘내적 타당성(internal validity)’과 ‘외적 타당성(external validity)’사이의 근본적인 긴장이 존재한다.

‘통제된 데이터’ 패러다임은 내적 타당성을 최우선으로 확보하려는 과학적 방법론이다. 내적 타당성이란 특정 연구 결과가 제3의 외적 변수가 아닌 오직 연구자의 의도된 처치(treatment)에 의해서만 비롯되었음을 보증하는 정도를 의미한다. 이는 결과의 반복 가능성(replicability)과 인과관계의 명확성을 중시하는 연역주의적(deductive) 연구 전통의 핵심 요건이다. M2VTS나 IEMOCAP과 같은 말뭉치들이 현실의 복잡성을 ‘혼란 변수(confounding variable)’로 간주하고 인위적으로 제거한 것은, 바로 이 내적 타당성을 확보하여 명확하고 반증 가능한(falsifiable) 지식을 얻으려는 목적 때문이었다(Pigeon & Vandendorpe, 1997; Busso et al., 2008).

반면, ‘생태학적 타당성’은 외적 타당성과 깊이 연관된 개념이다. 외적 타당성이란 특정 연구의 결과를 다른 시간, 장소, 그리고 다른 연구 대상에게까지 일반화할 수 있는 정도를 의미한다. 이는 귀납주의적(inductive) 연구 전통에서 특히 중시하는 가치로, 연구의 궁극적인 목표가 현실 세계에 대한 기술(description)과 설명에 있다고 보기 때문이다. FOLK나 CEJC와 같은 말뭉치들이 특정 가설 검증보다는 현실의 다양성을 최대한 반영하여 데이터를 수집한 것은, 연구 결과의 일반화 가능성, 즉 외적 타당성을 높이려는 의도였다(Schmidt, 2016; Koiso et al., 2018).

이 두 가지 타당성은 종종 하나를 얻으면 다른 하나를 잃게 되는 교환(trade-off) 관계에 놓인다. 내적 타당성을 높이기 위해 변수를 엄격하게 통제할수록 연구 환경은 현실과 멀어지게 되어 외적 타당성이 낮아질 수 있다. 반대로 외적 타당성을 높이기 위해 현실의 복잡성을 그대로 수용하면, 결과에 영향을 미치는 수많은 변수들을 통제할 수 없어 내적 타당성이 위협받게 된다.

결론적으로, 멀티모달 말뭉치 분야의 발전 과정은 이 두 가지 타당성 사이의 우선순위를 어떻게 설정할 것인가에 대한 연구 공동체의 지속적인 방법론적 고찰의 역사라고 할 수 있다.

3.4. 자연성 확보를 위한 방법론적 고찰

그러나 모든 과학적 질문이 명확한 인과관계 규명으로 환원될 수 있는 것은

아니다. 특히 인간의 사회적 행위나 언어와 같이 맥락 의존성이 높은 현상을 연구하는 분야에서는, 변수를 통제하는 과정 자체가 연구 대상을 왜곡하고 본질을 훼손할 수 있다는 비판이 꾸준히 제기되어 왔다. 이러한 배경에서 '통제'의 대안적 가치로 부상한 것이 바로 '자연성(naturalness)'이며, 이는 연구 결과의 '외적 타당성(external validity)', 특히 '생태학적 타당성(ecological validity)'과 깊이 연관된다.

'자연성'은 데이터가 연구자의 개입이나 인위적인 설정 없이, 실제 생활 세계에서 자발적으로 발생한 것인지를 따지는 데이터의 속성을 의미한다. 이는 연구의 목표가 인과관계의 증명이 아니라, 현실 세계에 대한 풍부하고 깊이 있는 기술(thick description)에 있다고 보는 귀납주의적(inductive) 연구 전통에 기반한다. 이러한 전통에서 데이터의 품질은 '통제' 수준이 아니라, 그것이 현실을 얼마나 충실하게 반영하는지에 따라 평가된다.

'자연성'의 기준을 확보하려는 노력은 여러 학문 분야에서 구체적인 방법론으로 발전했다. 예를 들어, 사회언어학 분야에서 윌리엄 라보프(William Labov)가 제기한 '관찰자 역설(Observer's Paradox)'은 연구자의 존재 자체가 연구 대상의 자연스러운 행동(특히 발화)을 왜곡하는 문제를 지적한다. 이를 극복하기 위해 연구자들은 장기간의 '참여 관찰'을 통해 연구 대상과 라포(rapport)를 형성하거나, 연구 목적을 숨긴 채 비개입적으로 데이터를 수집하는 등, 연구자의 영향을 최소화하기 위한 정교한 기법들을 고안해왔다. 또한 코퍼스 언어학에서는 수집된 데이터가 특정 장르나 화자에 편중되지 않고 연구하고자 하는 언어 공동체를 균형 있게 대표해야 한다는 '대표성(representativeness)'과 '균형성(balance)'을 '자연 데이터'의 핵심적인 품질 기준으로 삼는다.

결론적으로 '자연성'과 그로부터 파생되는 '생태학적 타당성'은, 통제를 통해 얻는 인과적 명료성 대신 현실 세계의 복잡성과 맥락성을 포착함으로써 연구의 일반화 가능성과 실질적 의미를 확보하려는 또 다른 과학적 이상을 대변한다.

자연성'을 확보하고 이를 통해 '생태학적 타당성'을 높이려는 시도는 여러 학문 분야에서 정교한 방법론적 성찰을 통해 발전해왔다. 이는 단순히 '현실을 있는 그대로 담아낸다'는 막연한 선언을 넘어, '무엇이 현실을 대표하는가' 그리고 '관찰 행위가 현실을 왜곡하지는 않는가'라는 근본적인 질문에 답하려는 노력이다.

3.4.1. 대표성의 문제와 브런즈윅(Brunswik)의 유산

'생태학적 타당성'이라는 개념의 지적 뿌리는 심리학자 에곤 브런즈윅(Egon

Brunswik)의 연구로 거슬러 올라간다. 그는 실험실 연구가 현실 세계에 대한 일반화 가능성을 가지려면, 연구에 사용된 자극이나 조건이 연구 대상이 실제 살아가는 환경의 특성을 대표적으로 표집(representative sampling)해야 한다고 주장했다(Brunswik, 1956). 즉, 연구의 타당성은 실험실 환경이 얼마나 현실과 '비슷하게 보이는지'가 아니라, 현실 세계의 다양한 조건들을 통계적으로 얼마나 잘 '대표'하는지에 달려있다는 것이다.

이러한 브런즈윅의 통찰은 코퍼스 언어학의 핵심 원리인 대표성(representativeness)과 균형성(balance)에 직접적으로 연결된다. 존 싱클레어(John Sinclair)와 같은 코퍼스 언어학의 선구자들은 '좋은' 자연 데이터 코퍼스가 단순히 실제 발화를 모아놓은 것이 아니라, 특정 언어 공동체의 언어 사용을 대표할 수 있도록 장르, 화자, 시대, 매체 등의 변인을 체계적으로 고려하여 구축되어야 한다고 강조했다(Sinclair, 1991). 이는 연구자가 자신의 직관에 의존하는 것이 아니라, 코퍼스라는 '대표적인 증거'에 기반하여 언어 현상을 기술해야 한다는 코퍼스 언어학의 경험주의적 신념을 반영한다.

3.4.2. 관찰의 문제와 라보프(Labov)의 역설

자연성을 확보하는 과정에서 가장 큰 난관은 '관찰자 역설(Observer's Paradox)'이다. 사회언어학자 윌리엄 라보프(William Labov)가 명료하게 정의한 이 문제는 "언어학 연구의 목표는 사람들이 체계적으로 관찰되고 있지 않을 때 어떻게 말하는지를 알아내는 것이지만, 우리는 오직 체계적인 관찰을 통해서만 그 데이터를 얻을 수 있다"는 딜레마를 가리킨다(Labov, 1972). 즉, 관찰 행위 자체가 데이터의 '자연성'을 훼손할 수 있다는 자기모순적 상황이다.

이 역설을 최소화하기 위해 사회언어학, 인류학, 대화분석 등의 분야에서는 다양한 방법론적 장치를 고안해왔다. 참여자가 녹음 사실을 잊을 만큼 감정적으로 몰입할 수 있는 주제(예: 죽음의 문턱에서 살아 돌아온 경험)에 대해 질문하는 기법, 혹은 연구자가 특정 공동체의 일원으로 장기간 거주하며 신뢰 관계를 형성한 뒤 자연스러운 상호작용을 기록하는 '참여 관찰(participant observation)' 기법 등이 대표적이다. 이는 연구자의 개입을 완전히 제거할 수는 없지만, 그 영향을 최소화하고 데이터의 진정성(authenticity)을 높이려는 정교한 노력이다.

3.4.3. 진정성과 맥락의 문제

더 나아가, 질적 연구와 민족지학(ethnography)에서는 데이터의 '자연성'을

넘어 '진정성(authenticity)'과 '맥락의 풍부함'을 강조한다. 이는 기록된 언어 데이터가 자발적으로 생성되었다는 사실을 넘어, 그 데이터가 발생한 문화적, 사회적 맥락 속에서 참여자들에게 어떤 의미를 갖는지를 이해하는 것이 중요하다는 입장이다. 인류학자 클리포드 기어츠(Clifford Geertz)가 말한 '두터운 기술(thick description)'은, 단지 행동을 묘사하는 것을 넘어 그 행동에 담긴 복잡한 의미의 층위를 해석해내는 과정을 의미한다(Geertz, 1973). 이 관점에서 본다면, 발화 내용만 있고 그 발화가 이루어진 사회적 관계나 문화적 배경에 대한 정보가 없는 데이터는 '자연적'일 수는 있어도 '진정한' 의미를 파악하기에는 근본적인 한계를 갖는다.

이처럼 브런즈윅의 통계적 대표성, 라보프의 관찰자 문제, 그리고 기어츠의 해석적 맥락성은 '자연성'이라는 개념이 단일한 기준으로 환원될 수 없는 다차원적인 구성물임을 보여준다. 오히려 이 기준들은 때로 서로 긴장 관계에 놓이기도 한다. 예를 들어, 브런즈윅의 제안에 따라 다양한 상황을 대표적으로 표집하려는 시도는 라보프가 지적한 관찰자 역설의 문제를 심화시킬 수 있으며, 기어츠가 요구하는 맥락에 대한 깊이 있는 이해는 광범위한 대표성을 확보하려는 목표와 상충될 수 있다.

결론적으로 '자연성'과 '생태학적 타당성'을 확보하는 것은 기술적인 차원을 넘어, 대표성에 대한 통계적 고민(브런즈윅), 관찰 행위에 대한 자기 성찰(라보프), 그리고 맥락에 대한 깊이 있는 이해(기어츠)를 요구하는 복합적인 학문적 실천이라 할 수 있다. 따라서 연구자가 어떤 기준을 우선시하는가는 해당 연구의 근본적인 인식론적 지향점과 직결되는 문제이다.

3.4.4. 자연성의 급진적 추구: 대화분석의 방법론

'자연성'이라는 가치를 가장 급진적이고 타협 없이 추구하는 연구 분야는 단연 대화분석(Conversation Analysis, CA)이다. 민족방법론(ethnomethodology)에 뿌리를 둔 대화분석은, 사회적 질서가 참여자들의 상호작용 속에서 순간순간 어떻게 구성되는지를 미시적으로 분석하는 것을 목표로 한다. 이러한 목표를 달성하기 위해, 대화분석은 연구 방법론의 모든 측면에서 '자연성'을 확보하기 위한 극도로 엄격한 기준을 적용한다.

대화분석의 제1원칙은 오직 '자연 발생적 상호작용(naturally occurring interaction)'만을 분석의 대상으로 삼는다는 것이다(Sacks, 1992). 여기서 '자연 발생적'이란, 연구자의 존재나 개입 여부와 상관없이 어차피 일어났을 상호작용

을 의미한다. 전화 통화, 가족의 저녁 식사 대화, 의사와 환자의 진료 상담 등, 연구를 위해 인위적으로 만들어진 상황이 아닌 실제 삶의 현장에서 벌어지는 상호작용이 바로 대화분석의 유일한 데이터이다. 이러한 원칙에 따라, 대화분석은 인터뷰, 설문, 실험 등 연구자가 특정 정보를 얻기 위해 상황에 개입하는 모든 형태의 데이터 수집을 근본적으로 배제한다. 왜냐하면 그러한 데이터는 참여자들이 현실 세계에서 실제로 사용하는 상호작용의 질서를 반영하는 것이 아니라, 연구자의 질문이나 실험 설계라는 인위적인 틀에 맞춰 재구성된 결과물이라고 보기 때문이다(ten Have, 2007).

이러한 엄격한 데이터 수집 원칙은 대화분석의 독특한 전사(transcription) 방법론으로 이어진다. 대화분석에서 전사는 단순히 녹음된 말을 글로 옮기는 부수적인 작업이 아니라, 분석의 핵심을 이루는 과정 그 자체이다. 특히 게일 제퍼슨(Gail Jefferson)에 의해 체계화된 전사 시스템은, 단어의 내용뿐만 아니라 상호작용의 질서를 구성하는 미세한 장치들을 포착하는 것을 목표로 한다(Jefferson, 1984). 여기에는 다음과 같은 요소들이 포함된다.

- 침묵의 측정: 1초 미만의 짧은 침묵(micropause)과 0.1초 단위로 측정된 침묵 시간을 정밀하게 표기한다.
- 발화의 중첩: 두 명 이상의 화자가 동시에 말하는 지점과 끝나는 지점을 대괄호 [ ]로 정확하게 표시한다.
- 음의 길이와 억양: 특정 음이 길게 발화되는 정도를 콜론 :으로, 말의 끝 억양이 올라가거나 내려가는 것을, ? .등 기호의 조합으로 표기한다.
- 목소리의 질: 웃음 섞인 목소리, 떨리는 목소리, 속삭이는 목소리 등 발화의 구체적인 전달 방식에 대한 정보를 담는다.

이처럼 정교한 전사 체계는, 언뜻 사소해 보이는 침묵의 길이, 발화의 중첩, 억양의 변화 등이 참여자들에게는 자신의 차례를 주장하거나, 상대방의 말에 동의 혹은 반대하거나, 대화의 주제를 전환하는 등 중요한 사회적 행위를 수행하는 자원(resource)이라는 대화분석의 핵심적인 발견에 기반한다(Sacks, Schegloff, & Jefferson, 1974).

결론적으로, 대화분석의 방법론은 '관찰자 역설'의 문제를 회피하는 것이 아니라, 자연 발생적 데이터에 대한 원칙과 미시적 전사라는 도구를 통해 정면으로 돌파하려는 시도라고 할 수 있다. 이는 연구자의 이론적 전제를 데이터에 부과하는 것이 아니라, 데이터 속에서 참여자들이 스스로 드러내는 질서를 발견하려는, 가장 철저하게 귀납적이고 경험적인 '자연성' 확보의 실천이다.

3.4.4.1. 관찰자 역설의 극복: 자연 발생적 데이터의 수집 전략

대화분석이 이처럼 엄격한 '자연 발생적 데이터' 원칙을 고수하기 때문에, 데이터 수집 단계에서부터 '관찰자 역설'의 문제를 극복하기 위한 체계적인 전략이 요구된다. 대화분석 연구자들이 활용하는 수집 전략은 연구자의 개입을 최소화하여 데이터의 자연성을 확보하려는 공통된 목표를 갖는다.

가장 이상적인 데이터는 연구자가 녹음 과정에 전혀 개입하지 않은 '발견된 데이터(found data)'이다. 대화분석 분야를 개척한 하비 색스(Harvey Sacks)의 초기 연구는 그가 직접 녹음한 것이 아니라, 자살예방센터에 보관되어 있던 상담 전화 녹음 테이프를 분석한 것이었다(Sacks, 1992). 이와 유사하게, 방송 토론, 뉴스 인터뷰, 법정 기록, 항공 관제 교신, 응급 구조 요청 전화 등은 연구자의 존재가 상호작용에 전혀 영향을 미치지 않은 순수한 자연 데이터의 원천이 된다.

그러나 연구 주제에 맞는 '발견된 데이터'를 항상 확보할 수 있는 것은 아니기 때문에, 연구자가 직접 녹음을 해야 하는 경우가 많다. 이 경우, 연구자의 영향을 최소화하기 위한 다양한 기법이 사용된다. 한 가지 방법은 연구 대상에게 녹음 장비를 제공하고, 연구자가 없는 상황에서 그들끼리의 상호작용(예: 가족의 식사 시간, 친구와의 통화)을 직접 녹음해달라고 요청하는 것이다. 이는 연구자가 물리적으로 존재하지 않음으로써 관찰자 효과를 줄이는 효과적인 전략이다(ten Have, 2007).

또 다른 핵심적인 전략은 기관 환경(institutional settings)의 상호작용을 활용하는 것이다. 병원, 콜센터, 학교, 상담소 등 많은 기관에서는 업무의 일환으로(예: 품질 관리, 교육, 법적 기록 등) 상호작용을 일상적으로 녹음한다. 연구자는 기관의 허가를 얻어 이러한 기록물에 접근함으로써, 참여자들이 연구가 아닌 자신의 본래 업무에 집중하는 동안 생성된 자연스러운 데이터를 확보할 수 있다. 이 경우, 참여자들은 녹음 사실을 인지하더라도 '연구 대상'으로서의 의식보다는 '의사', '상담원', '교사'로서의 역할 수행에 더 몰두하기 때문에 데이터의 자연성이 높게 유지된다.

이러한 전략들은 모두 연구자의 발자국을 데이터에서 최대한 지우고, "어차피 일어났을" 상호작용에 가장 근접한 기록을 확보하려는 대화분석의 근본적인 방법론적 지향점을 보여준다.

3.4.4.2. 카메라의 존재가 주는 불편함: 대화분석의 '자연성'에 대한 변호

대화분석의 방법론에 대해 제기되는 가장 근본적인 비판은 녹음기나 카메라

의 존재 자체가 상호작용의 '자연성'을 필연적으로 훼손한다는 것이다. 즉 대화분석에서 목표로 하는 자연성은 기록하는 순간 근원적 달성할 수 없는 목표가 된다는 것이다. 그러나 대화분석 연구자들은 이러한 비판에 대해 몇 가지 중요한 이론적, 경험적 답변을 제시한다.

1. 상호작용의 요구가 관찰자 인식을 압도한다: 대화분석의 가장 핵심적인 답변은, 실제 상호작용의 순간순간에 발생하는 과업(예: 말차례 교대, 오해의 수정, 이야기 전달)이 너무나 복잡하고 즉각적인 인지적 자원을 요구하기 때문에, 참여자들은 자신이 녹화되고 있다는 사실을 지속적으로 의식하기 어렵다는 것이다. 즉, '사회적 상호작용의 압력'이 '연구 참여자로서의 자기 인식'을 빠르게 압도하고 무력화시킨다.
2. 데이터 속 증거: 참여자의 지향(Orientation) 문제: 대화분석은 이 주장을 추상적으로만 내세우지 않고, 데이터 속에서 경험적으로 증명할 수 있다고 본다. 만약 참여자가 카메라나 녹음기를 의식하고 있다면, 그들의 발화나 행동 속에서 그러한 인식을 드러내는 증거, 즉 카메라를 향한 '지향(orientation)'이 나타날 것이다(예: "이 말은 녹음되면 안 되는데"와 같은 발언, 카메라를 의식하는 시선 처리 등). 그러나 방대한 실제 데이터 분석을 통해 밝혀진 것은, 참여자들이 압도적으로 카메라가 아닌 '서로'를 지향하며 상호작용에 몰두한다는 사실이다.
3. 환경의 일부가 된 기록 장비: 특히 기관 환경(institutional settings) 연구에서, 기록 장비는 연구자를 위해 외부에서 이식된 이물질이 아니라, 원래부터 그 환경의 일부인 경우가 많다. 콜센터의 통화 녹음이나 법정의 속기록처럼, 기록은 참여자들이 수행하는 과업의 자연스러운 일부이며, 따라서 그 존재가 상호작용을 특별히 왜곡한다고 보기 어렵다.
4. 내적 상태의 배제: 마지막으로, 대화분석은 참여자가 '심리적으로 자연스럽다고 느끼는지'와 같은 내적 상태를 분석의 대상으로 삼지 않는다. 대화분석의 관심은 오직 참여자들이 상호작용 속에서 공개적으로 드러내고, 서로가 반응하는 행동에 있다. 만약 참여자가 카메라의 존재를 상호작용의 문제로 삼지 않는다면, 그것은 분석적으로도 유의미한 변수가 아니라고 간주한다.

3.4.4.3. 대화분석의 실천적 해법: 수집 전략과 연구 사례

대화분석이 '자연 발생적 데이터'를 확보하기 위해 사용하는 수집 전략들은 단순한 이론적 원칙을 넘어, 구체적인 연구 사례들을 통해 그 실효성이 입증되

어 왔다.

가장 이상적인 데이터 수집 방식인 '발견된 데이터(found data)' 활용의 대표적인 사례는 단연 하비 색스(Harvey Sacks)의 초기 연구다. 그는 로스앤젤레스 자살예방센터에 보관되어 있던 상담 전화 녹음 기록을 분석하여, "전화를 거는 사람은 자신의 이름을 먼저 말하지 않는다"와 같은 일상 대화의 암묵적인 규칙들을 최초로 발견했다(Sacks, 1992). 또한, 클레이먼과 헤리티지(Clayman & Heritage, 2002)는 미국 대통령과 기자들 사이의 기자회견 녹화 자료를 분석하여, 기자들이 어떻게 질문의 형식을 통해 대통령에게 공격적이거나 비판적인 입장을 드러내는지를 미시적으로 분석했다. 이 연구들은 방송이나 기관의 업무 절차에 따라 자연스럽게 기록된 데이터를 활용하여, 연구자의 개입을 원천적으로 차단한 사례이다.

연구자가 직접 녹음을 추진해야 할 경우, 연구자 개입을 최소화하는 전략이 사용된다. 이 접근법의 대표적인 사례는 찰스 굿윈과 마저리 하네스 굿윈(Charles Goodwin & Marjorie Harness Goodwin)부부의 장기간에 걸친 민족지학적 비디오 연구다. 이들은 친구나 이웃의 가정에 비디오카메라를 설치하고 장시간 녹화하여, 가족의 식사 시간이나 아이들의 놀이와 같은 지극히 일상적인 상호작용을 포착했다(Goodwin, 1990). 이 방대한 비디오 데이터를 통해 그들은 발화, 시선, 몸짓, 그리고 주변 환경의 사물이 어떻게 통합되어 하나의 사회적 행위를 구성하는지를 분석했으며, 이는 멀티모달 상호작용 연구의 초석이 되었다.

기관 환경(institutional settings)의 일상적인 기록 절차를 활용하는 것 또한 핵심적인 전략이다. 안시 페래퀼래(Anssi Peräkylä, 1995)는 AIDS 상담 클리닉에서 의사와 환자 간의 상담 과정을 녹화한 데이터를 분석했다. 이 상황에서 녹화는 연구 목적이 아니라 의사 훈련의 일환으로 이루어졌기 때문에, 참여자들은 연구가 아닌 상담이라는 본연의 제도적 과업에 몰두할 수 있었다. 페래퀼래는 이 데이터를 통해 의사와 환자가 죽음이라는 극도로 민감한 주제를 어떻게 회피하거나 혹은 조심스럽게 꺼내어 이야기하는지를 세밀하게 분석하여, 상담 상호작용의 구조를 밝혀냈다.

이러한 연구 사례들은 대화분석의 '자연성' 확보 원칙이 단순한 이론적 이상에 그치는 것이 아니라, 연구 질문에 맞춰 다양한 수집 전략을 창의적으로 적용하는 구체적이고 실천적인 방법론임을 명확히 보여준다.

4. 패러다임의 렌즈로 본 음성언어 말뭉치 재평가

앞선 3장에서는 멀티모달 말뭉치 구축의 기저에 있는 두 가지 상충하는 패러다임, 즉 내적 타당성을 추구하는 '통제'와 외적 타당성을 지향하는 '자연성'에 대해 심층적으로 고찰했다. 본 장에서는 이 이론적 틀을 분석의 렌즈로 삼아, 2장에서 개관한 주요 음성언어 말뭉치들을 비판적으로 재평가하고자 한다. 각 말뭉치가 두 패러다임의 긴장 관계 속에서 어떤 방법론적 선택을 했으며, 그 선택이 어떤 학문적 가치를 창출하고 또 어떤 한계를 내포하게 되었는지를 분석한다.

4.1. 내적 타당성의 극대화: 통제된 데이터의 사례

멀티모달 말뭉치 발전의 초기 단계는 공학적 문제 해결이라는 뚜렷한 목표 아래, 연구의 내적 타당성(internal validity)을 극대화하려는 시도로 요약될 수 있다. 내적 타당성이란 특정 연구 결과가 제3의 외적 변수가 아닌 오직 연구자의 의도된 처치(treatment)에 의해서만 비롯되었음을 보증하는 정도를 의미한다. 1세대 말뭉치인 M2VTS는 이러한 통제 패러다임의 전형을 보여준다(Pigeon & Vandendorpe, 1997). '음성과 얼굴 정보의 융합이 개인 식별의 정확도를 향상시키는가?'라는 명확한 인과적 질문에 답하기 위해, 연구팀은 배경, 조명, 발화 내용 등 결과에 영향을 미칠 수 있는 모든 잠재적 혼란 변수를 의도적으로 제거했다. 이처럼 현실을 극도로 단순화한 실험 환경은, 알고리즘 성능의 변화가 오직 '멀티모달 융합'이라는 변수에 의해서만 발생했음을 명확히 증명하기 위한 필수적인 선택이었다. 이로써 M2VTS는 높은 내적 타당성을 확보했지만, 그 데이터가 실제 생활 속 의사소통 상황을 거의 반영하지 못한다는 점에서 외적 타당성, 특히 생태학적 타당성은 현저히 낮을 수밖에 없었다.

'감성'이라는 더 복잡하고 주관적인 영역을 다루기 시작한 3세대 말뭉치 IEMOCAP역시 통제 패러다임의 정교한 적용 사례로 볼 수 있다(Busso et al., 2008). 연구팀은 '특정 감정은 특정 비언어적 신호와 체계적으로 연결된다'는 가설을 탐구하기 위해, '자연스러운 감정'을 관찰하는 대신 '전문 배우의 감정 연기'를 데이터로 수집하는 방식을 택했다. 고정밀 모션 캡처 기술을 동원하여 배우의 미세한 신체 움직임을 정량화한 것은, '감정'이라는 모호한 개념에 대해 신뢰도 높은 정답(ground-truth)을 부여하고 분석의 정밀도를 높이려는 시도였다. 이 경우, 측정된 비언어적 신호가 '의도된 감정 범주'와 명확히 연결된다는 점에서 높은 내적 타당성을 확보했다고 할 수 있다. 그러나 이 역시 연기된 감정

이 실제 생활에서 자발적으로 발생하는 감정과 동일하다고 보장할 수 없다는 근본적인 한계를 안고 있었다(Busso et al., 2008).

결론적으로, 통제 패러다임에 기반한 M2VTS와 IEMOCAP과 같은 말뭉치들은 해당 분야의 발전에 필수적인 기여를 했다. 이들은 현실을 의도적으로 단순화함으로써, '멀티모달 융합은 가능한가?' 혹은 '감정을 계산적으로 모델링할 수 있는가?'와 같은 근본적인 질문에 대한 명료한 '개념 증명(proof of concept)'을 제공했다. 이러한 말뭉치들은 특정 인과적 질문에 답하는 데에는 매우 효과적이었으나, 그 결과가 현실의 복잡한 상호작용에까지 일반화되기는 어렵다는 명백한 한계를 가졌다. 바로 이 한계에 대한 인식이, 이후 연구자들이 외적 타당성을 중시하는 자연성 패러다임으로 눈을 돌리게 만드는 결정적인 계기가 되었다.

4.2. 외적 타당성의 추구: 자연성 패러다임의 사례

'통제된 데이터' 패러다임이 가진 일반화의 한계에 대한 인식이 깊어지면서, 2010년대 연구자들은 연구의 무게중심을 내적 타당성에서 외적 타당성(external validity), 특히 생태학적 타당성으로 옮기기 시작했다. 이는 현실 세계의 복잡한 상호작용을 있는 그대로 기술(description)하고 이해하는 것을 연구의 우선적인 목표로 삼는 패러다임의 전환을 의미한다. 이러한 말뭉치들은 연구 결과가 실험실을 넘어 실제 생활 세계에까지 적용될 수 있는 가능성을 높였지만, 반대로 수많은 변수가 통제되지 않은 상태이므로 특정 현상에 대한 명확한 인과관계를 주장하기는 어려운 한계를 갖는다.

이러한 자연성 패러다임을 대표하는 FOLK와 CEJC는 중요한 언어학적 성과물로서, 앞서 3장에서 논의된 자연성 확보를 위한 엄밀한 기준들을 충족시키려는 노력을 보여준다. 특히 이들 말뭉치의 데이터 수집 방식은, 연구자의 개입을 최소화하고 참여자의 실제 삶 속에서 발화가 생성되도록 한다는 점에서 대화분석(Conversation Analysis)이 추구하는 엄격한 데이터 수집 원칙과 그 방향을 같이 한다. FOLK는 다양한 사회적 맥락을 체계적으로 수집하여 데이터의 대표성을 높이고자 했으며(Schmidt, 2016), CEJC는 '개인 기반 수집' 방식을 통해 관찰자 효과를 최소화하고 데이터의 진정성을 확보하려는 방법론적 혁신을 보여주었다(Koiso et al., 2018). 이는 이들 말뭉치가 단순한 데이터 모음이 아니라, 언어 현상을 그것이 발화되는 실제 사회적 맥락 속에서 이해하려는 언어학의 오랜 목표를 충실히 구현한 결과물임을 의미한다.

반면, 5세대의 CANDOR말뭉치는 방대한 규모와 풍부한 심리 데이터라는

가치에도 불구하고, 언어학적 관점에서는 중요한 방법론적 후퇴를 보여준다(Reece et al., 2023). 첫째, CANDOR는 막대한 규모를 확보하는 과정에서 '정밀성(precision)'을 희생했다. 데이터 전사를 자동 음성 인식(ASR)에 의존함으로써(Reece et al., 2023), 대화분석이나 담화분석에서 상호작용의 핵심 자원으로 간주하는 미세한 발화 중첩, 침묵, 억양 변화 등을 포착하지 못하는 명백한 한계를 갖는다. 둘째, 그리고 더 근본적으로, CANDOR는 '자연성'자체에 대한 기준을 완화했다. 이 말뭉치의 데이터는 연구를 위해 인위적으로 짝지어진 초면의 참가자들사이의 대화로 구성된다. 이는 가족이나 친구 사이의 대화처럼 참여자들의 실제 삶 속에서 자연스럽게 발생한 상호작용이 아니라, '연구에 참여한다'는 특수한 목적을 가진 인위적인 사회적 상황이다. 따라서 CANDOR의 사례는 자연성 패러다임 내부에서도 '규모'와 '정밀성', 그리고 '규모'와 '자연성' 사이의 새로운 상충 관계가 존재함을 보여준다.

4.3. 긴장의 교차점: 복합적 패러다임의 사례

'통제된 데이터'와 '자연성' 패러다임이 항상 명확하게 분리되는 것은 아니다. 몇몇 중요한 말뭉치들은 두 패러다임의 경계에서, 혹은 의도적으로 두 패러다임을 절충하려는 시도로서 구축되었다. 그러나 이러한 시도들 역시 두 패러다임 사이의 근본적인 긴장을 완전히 해소하지는 못했으며, 오히려 그 긴장이 어떻게 다른 방식으로 나타나는지를 보여주는 흥미로운 사례가 된다. 2세대 말뭉치인 AMI는 이러한 절충을 위한 초기 시도로 볼 수 있다. AMI는 '시나리오 기반 유도' 방식을 통해, 참가자들이 특정 과업(TV 리모컨 설계)을 수행하게 함으로써 상호작용에 일정한 구조를 부여했다. 이는 '통제' 패러다임의 요소를 차용한 것이지만, 그 목표는 통제된 환경 속에서 최대한 자연스러운 회의 담화를 유도하려는, 즉 '자연성' 패러다임의 가치를 일부 구현하려는 것이었다(Carletta et al., 2005). 하지만 이 방법론은 근본적으로 한계를 가질 수밖에 없다. 참가자들은 실제 자신의 업무가 아닌 주어진 역할을 연기하는 것이므로, 이들의 상호작용은 진정한 의미에서 '자연 발생적'이라고 보기 어렵다. 결국 AMI는 두 패러다임의 긴장을 해소하기보다는, 그 긴장 속에서 타협점을 찾으려 한 과도기적 시도로 평가할 수 있다.

5세대의 MISP챌린지 말뭉치는 훨씬 더 복잡하고 현대적인 방식으로 두 패러다임의 긴장을 드러낸다. MISP의 데이터는 TV 소음과 발화 겹침이 난무하는 실제 가정 환경에서 수집되었다는 점에서, 표면적으로는 4세대 말뭉치들이 추구했

던 극도의 '자연성(naturalness)' 혹은 '생태학적 현실성(ecological realism)'을 담고 있다(Wang et al., 2023). 그러나 이 말뭉치의 구축 목적은 상호작용의 의미를 이해하거나 언어 현상을 기술하려는 것이 아니다. 오히려 이 현실의 복잡성은 '알고리즘이 극복해야 할 장애물'로서 기능하며, 말뭉치의 유일한 목표는 '화자 분할 및 인식'이라는 명확한 공학적 과제에 대한 알고리즘의 성능을 측정하고 순위를 매기는 것이다(Wang et al., 2023). 즉, MISP는 방법론적으로 가장 '자연스러운' 데이터를 수집하여, 목적론적으로는 가장 '통제된' 방식으로 소비한다.

AMI와 MISP의 사례는 '통제'와 '자연성' 사이의 긴장이 단순히 과거에서 현재로 이어지는 선형적 발전 과정 속에서 해소되는 문제가 아님을 명확히 보여준다. AMI의 시도는 어설픈 타협으로 인해 두 패러다임의 장점을 모두 놓칠 수 있는 위험을 보여주었으며, MISP의 사례는 가장 '자연스러운' 데이터조차도 연구의 목적에 따라 그 의미가 완전히 탈각되고 순수한 공학적 '입력값'으로 전락할 수 있음을 보여준다. 이는 결국 어떤 말뭉치든 그 가치는 절대적으로 평가될 수 없으며, 오직 3장에서 논의한 이론적 패러다임과 연구자가 던지는 특정 질문과의 관계 속에서만 규정될 수 있음을 시사한다.

4.4. 긴장 관계의 동학: 음성언어 말뭉치 분석의 함의

본 장의 분석을 통해, 멀티모달 말뭉치의 발전사가 내적 타당성을 추구하는 '통제' 패러다임과 외적 타당성을 지향하는 '자연성' 패러다임 사이의 지속적인 긴장 관계 속에서 이루어져 왔음을 확인했다. M2VTS부터 FOLK/CEJC에 이르기까지, 모든 말뭉치는 이 두 가치 사이에서 특정 방법론을 선택한 결과물이다. 그렇다면 전례 없는 규모를 자랑하는 5세대 말뭉치들은 이 오랜 긴장을 해소했는가? 본 장의 분석은 오히려 그 반대를 시사한다. 즉, 규모의 성장은 이 긴장을 해소하는 것이 아니라, 더 복잡하고 새로운 방식으로 드러나게 만든다.

CANDOR의 사례는 규모를 확보하기 위해 '자연성'의 다른 측면이 어떻게 희생될 수 있는지를 보여준다. CANDOR는 분명 자발적인 대화를 담고 있지만, 그 대화는 연구를 위해 인위적으로 짜지어진 초면의 참가자들 사이에서 이루어진다. 이는 가족, 친구, 동료 사이에서 발생하는, 참여자들의 실제 삶에 뿌리내린 상호작용이 아니다. 결국 CANDOR는 방대한 데이터를 얻기 위해, 대화 내용의 자발성이라는 가치는 유지하되 사회적 관계의 자연성이라는 가치는 포기하는, 일종의 '통제된 자연성'을 선택한 셈이다. 이는 규모의 증가라는 양적 팽창이 데이터의 질적 측면, 특히 생태학적 타당성의 근간이 되는 사회적 진정성

(social authenticity)의 문제를 해결해주지 못함을 보여준다.

MISP는 이 긴장이 훨씬 더 근본적인 차원에서 작동함을 보여준다. MISP는 TV 소음과 발화 겹침이 가득한, 환경적으로는 극도로 '자연스러운' 데이터를 수집한다. 그러나 이 말뭉치의 목적은 언어 현상을 기술하거나 상호작용의 의미를 이해하는 데 있지 않다. 오히려 이 현실의 복잡성은 '알고리즘이 극복해야 할 장애물'로서만 기능하며, 말뭉치의 유일한 목표는 '알고리즘 성능 측정'이라는 명확한 공학적 과제를 수행하는 것이다. 즉, MISP는 언어 데이터가 갖는 상호작용적 의미를 완전히 제거하고 순수한 신호(signal)로 환원함으로써, 말뭉치가 가져야 할 근본적인 성격 자체를 공학적 목표라는 궁극적인 '통제'에 종속시킨다.

결론적으로, 본 장의 분석은 '통제'와 '자연성' 사이의 긴장이 멀티모달 말뭉치 설계에 있어 피할 수 없는 근본적인 조건임을 재확인시켜 준다. 규모의 성장은 이 문제를 해결하는 것이 아니라, '어떤 종류의 자연성을 희생하여 어떤 종류의 통제를 할 것인가'라는 더 복잡한 질문을 제기할 뿐이다. 본 장에서 도출된 이 분석틀은, 비단 음성언어 말뭉치에만 국한되지 않는다. 이 분석틀이 시각 채널이 중심이 되는 수어 말뭉치라는 특수한 사례를 분석하는 데에도 유용한 도구가 될 수 있음을 다음 장에서 논의하고자 한다.

5. 수어 말뭉치의 분석

지금까지의 논의는 음성-청각 채널을 중심으로 하는 음성언어 말뭉치에 집중되었다. 본 장에서는 앞서 3장에서 도출된 분석틀, 즉 '통제'와 '자연성' 사이의 긴장 관계라는 렌즈를 통해, 시각-운동 채널을 제1의 언어 채널로 사용하는 수어 말뭉치라는 특수한 사례를 분석하고자 한다. 수어 말뭉치는 음성언어 말뭉치와는 다른 고유한 방법론적 과제를 안고 있으며, 이는 두 패러다임의 긴장이 얼마나 복합적인 양상으로 나타날 수 있는지를 보여주는 중요한 시금석이 된다.

5.1. 시각 언어의 특수성: 수어 말뭉치의 방법론적 과제

수어 말뭉치는 음성언어 말뭉치와 근본적으로 다른 여러 특성을 가지며, 이는 데이터 수집에서부터 분석에 이르기까지 전 과정에 걸쳐 새로운 방법론적 과제를 제기한다. 첫째, 언어 정보가 생성되고 수용되는 주된 채널이 음성-청각에서 시각-운동 채널로 전환된다. 이는 단순히 녹음기 대신 카메라가 필요하다

는 기술적 차원을 넘어선다. 화자의 미세한 손가락 움직임부터 상체의 위치 변화까지, 언어 정보를 담고 있는 모든 시각적 요소를 누락 없이 포착하기 위해 다각도의 카메라 설치나 정교한 조명 설계가 필수적이다.

둘째, 수어 문법은 동시성(simultaneity)이라는 중요한 특징을 갖는다. 음성언어가 주로 선형적인 음소의 배열로 구성되는 것과 달리, 수어에서는 손의 모양과 위치, 움직임(수동 신호)과 함께 얼굴 표정, 눈썹, 입 모양, 상체의 기울기 등 다양한 비수지 신호(Non-Manual Signals, NMS)가 동시에 결합하여 하나의 문법적 단위를 형성한다. 이는 선형적인 텍스트 전사를 넘어, 다층적이고 동시적인 정보를 담아낼 수 있는 정교한 주석(annotation) 체계를 요구하며, 이는 수어 말뭉치 구축의 핵심적인 난제가 된다.

셋째, 수어가 시각 언어임에도 불구하고, 상호작용을 총체적으로 이해하기 위해서는 소리 정보의 수집이 여전히 중요한 과제가 된다. 수어 화자가 청인과 상호작용하는 경우, 청인의 음성 발화는 대화의 절반을 차지하는 핵심 데이터이다. 또한 수어 화자 자신도 웃음이나 한숨과 같은 의미 있는 발성을 하거나, 음성언어의 단어를 입 모양으로 함께 발화하는 구화(mouthing)를 사용한다. 나아가, 대화가 이루어지는 환경음은 상호작용의 흐름에 영향을 미치는 중요한 맥락적 단서가 될 수 있다. 따라서 온전한 상호작용 말뭉치를 구축하기 위해서는 시각 정보와 함께 음향 정보까지 포괄적으로 수집해야 한다는 새로운 요구에 직면하게 된다.

5.2. 해외 주요 수어 말뭉치의 경향과 사례

수어학 연구가 발전함에 따라, 다양한 국가에서 대규모 수어 말뭉치를 구축하려는 노력이 활발하게 이루어져 왔다. 이러한 프로젝트들은 각기 다른 규모와 구체적인 목표를 갖지만, 데이터 수집 방법론과 구성에 있어 몇 가지 공통된 경향을 보여준다.

이러한 대규모 프로젝트의 가장 대표적인 사례는 독일수어 말뭉치(DGS-Korpus)이다. 15년에 걸친 장기 프로젝트로 기획된 이 말뭉치는, 330명 이상의 농인 참여자로부터 560시간 이상의 방대한 비디오 데이터를 구축했다(Jahn et al., 2018; Prillwitz et al., 2008). DGS-Korpus의 가장 큰 특징은 독일수어의 언어적 다양성을 포괄적으로 기록하기 위해, 자유로운 대화뿐만 아니라 특정 어휘나 문법 구조를 유도하기 위한 이야기 재화(story retelling), 농담, 경험담 보고 등 18가지에 달하는 다양한 유도 과제(elicitation task)를 병행했다는 점이다(Jahn et al.,

2018). 이는 수어의 실제 사용 양상을 기록하려는 목적과, 특정 언어 단위의 비교 분석을 용이하게 하려는 목적을 동시에 추구하는 현대 수어 말뭉치의 경향을 잘 보여준다.

이러한 경향은 다른 주요 수어 말뭉치 프로젝트에서도 유사하게 발견된다. 호주수어 말뭉치(Auslan Corpus)는 100명의 농인 모어화자 및 준모어 화자를 대상으로 인터뷰, 이야기 발화, 시각 자극 묘사, 자유 대화 등 다양한 종류의 데이터를 수집했다(Johnston, 2008, 2010). 네덜란드수어 말뭉치(Corpus NGT)역시 다양한 연령대와 지역 배경을 가진 100명의 모어 화자로부터 여러 장르의 데이터를 수집했으며(Crasborn & Zwitserlood, 2008), 특히 화자의 상반신 정면 모습과 함께 머리 위에서 촬영하는 상면 시점(top view)을 포함한 다중 시점 녹화를 진행하여 수어의 3차원적 공간 사용을 기록하고자 했다. 미국수어(ASL) 말뭉치를 구축하는 NCSLGR(National Center for Sign Language and Gesture Resources) 역시 정면, 측면, 그리고 얼굴 클로즈업 등 다중 동기화 비디오를 제공하는 것을 표준으로 삼고 있다(Neidle, 2007). 이들 사례를 통해, 많은 해외 수어 말뭉치들이 언어 현상의 자연스러운 사용을 포착하려는 목적과, 특정 언어 단위를 비교 분석하기 위해 통제된 환경에서 데이터를 유도하려는 목적을 동시에 추구하며, 이를 위해 자유 대화와 통제된 과제를 혼합하는 전략을 취하는 일반적인 경향이 있음을 확인할 수 있다.

5.3. ‘통제된 데이터’와 ‘자연성’의 교차점으로서의 수어 말뭉치

앞선 사례들을 통해 살펴본 것처럼, 수어 말뭉치는 본고의 핵심 분석틀인 ‘통제된 데이터’와 ‘자연성’ 패러다임의 긴장이 첨예하면서도 복합적으로 드러나는 영역이다. 한편으로 수어 말뭉치는 농인 커뮤니티의 ‘자연 언어(수어)’를 대상으로 한다는 점에서 외적 타당성을 극단적으로 추구한다. 그러나 다른 한편으로는, 그 자연 언어를 데이터로 만드는 과정에서 음성언어 말뭉치보다 훨씬 더 강력하고 다층적인 ‘통제’가 개입될 수밖에 없는 근본적인 딜레마를 안고 있다.

그 첫 번째 층위는 데이터 수집 환경의 물리적 제약에서 발생한다. 지적된 바와 같이, 수어는 3차원 공간에서 손의 정교한 움직임, 얼굴 표정, 상체의 이동 등 다양한 시각 정보를 통해 의미를 전달한다. 이러한 언어적 정보를 누락 없이 포착하기 위해서는 단일 마이크로 수음이 가능한 음성언어와 달리, 정면, 측면, 상면 등 다각도의 고화질 카메라와 균일한 조명 장비가 필수적이다(Crasborn & Zwitserlood, 2008; Neidle, 2007). 이러한 장비의 설치는 필연적으로 일상적인

대화 환경을 인위적인 스튜디오나 실험실 환경으로 변화시킨다. 즉, 데이터의 정밀성을 확보하려는 기술적 '통제' 수준이 높아질수록, 관찰자 역설의 문제 또한 심화되어 상호작용의 '자연성'이 위협받는 정도가 더 커지는 것이다.

두 번째 층위는 데이터의 종류를 선택하는 과정에서 나타난다. 5.2절에서 살펴본 바와 같이, DGS-Korpus는 자유 대화와 함께 특정 어휘나 문법 구조를 수집하기 위한 '유도 과제(elicitation task)'를 적극적으로 활용한다(Jahn et al., 2018). 이는 언어 현상의 자연스러운 사용을 포착하려는 목적과, 특정 언어 단위를 비교 분석하기 용이한 데이터를 확보하려는 목적 사이에서 이루어지는 의식적인 타협이다. 즉, '자연성' 패러다임의 목표를 달성하기 위해 '통제' 패러다임의 방법론을 적극적으로 차용하는 것이다.

마지막으로, 주석 과정 역시 일종의 분석적 '통제'라고 할 수 있다. 수어에서 동시에 발현되는 수동 신호와 비수지 신호의 복합적인 정보를 ELAN과 같은 주석 도구의 개별적인 층렬에 선형적으로 나누어 기록하는 행위 자체가, 연속적이고 총체적인 시각 정보를 분석 가능한 불연속적 단위로 재구성하는 과정이다. 결론적으로 수어 말뭉치는 '자연성'을 지향하면서도 그 구현을 위해 기술적, 방법론적, 분석적 차원에서 '통제'에 깊이 의존할 수밖에 없는, 두 패러다임의 긴장이 가장 역동적으로 교차하는 흥미로운 연구 대상이라 할 수 있다.

5.4. '통제된 데이터'와 '자연성'의 렌즈로 본 해외 말뭉치 평가

앞선 절들에서 개관한 해외 주요 수어 말뭉치들은 본고의 핵심 분석틀인 '통제된 데이터'와 '자연성' 패러다임의 긴장 관계가 실제 말뭉치 설계 과정에서 어떻게 작동하는지를 보여주는 구체적인 사례다. 이들 말뭉치는 어느 한 패러다임에 온전히 속하기보다는, 연구 목적을 달성하기 위해 두 패러다임의 요소를 의식적으로 선택하고 절충하는 양상을 공통으로 보여준다.

첫째, 데이터 수집 방식의 차원에서 명시적인 방법론적 절충이 이루어진다. 예를 들어, DGS-Korpus는 농인들의 자연스러운 대화를 수집하면서도, 특정 언어 현상(분류사, 일치동사 등)의 비교 분석을 용이하게 하기 위해 이야기 재화(story retelling)를 포함한 18가지에 달하는 유도 과제를 병행했다(Jahn et al., 2018; Prillwitz et al., 2008). Auslan Corpus 역시 자유 대화와 인터뷰, 특정 시각 자극에 대한 묘사 과제를 혼합하여 데이터를 수집했다(Johnston, 2008, 2010). 이는 '자연성'을 통해 외적 타당성을 확보하려는 노력과 동시에, 특정 변수를 비교 가능한 형태로 수집하려는 '통제'의 필요성을 인정한 전략적인 선택이다.

이는 두 패러다임의 가치 사이에서 이루어지는 의도적인 타협이라 평가할 수 있다.

둘째, 데이터의 정밀도를 확보하기 위한 기술적 통제는 필연적으로 녹화 환경의 자연성을 저해하는 결과를 낳는다. 5.1절에서 논의했듯이, 수어의 3차원적 공간 사용과 미세한 비수지 신호를 포착하기 위해서는 다중 카메라 시스템이 필수적이다. Corpus NGT나 NCSLGR과 같은 프로젝트들이 정면, 상면, 측면 등 여러 각도에서 동기화된 영상을 촬영하는 것은, 데이터의 정밀성이라는 가치를 극대화하기 위한 명백한 '통제' 장치다(Crasborn & Zwitserlood, 2008; Neidle, 2007). 그러나 이러한 기술적 요구는 촬영 환경을 일상적인 공간이 아닌 스튜디오나 실험실과 유사한 공간으로 만들며, 이는 음성언어 말뭉치 수집에 비해 관찰자 역설의 문제를 더욱 심화시키는 요인으로 작용한다.

5.5. 수어 말뭉치 연구의 근본적 딜레마와 그 해법

앞선 평가에서 확인했듯이, 수어 말뭉치 구축은 '통제'와 '자연성' 사이의 여러 긴장 관계에 놓여있다. 이 절에서는 한 걸음 더 들어가, 수어 말뭉치 연구자들이 이러한 긴장 속에서 '자연성'을 확보하기 위해 직면하는 두 가지 근본적인 딜레마와, 이에 대한 학계의 이론적, 실천적 해법을 심층적으로 다룬다.

첫 번째 딜레마는 '기술적 정밀성'과 '관찰자 효과' 사이의 상충 관계, 즉 카메라의 침입성 문제이다. 음성언어와 달리 수어의 문법 정보는 손의 정교한 움직임뿐 아니라 얼굴 표정, 몸의 기울기 등 다양한 비수지 신호에 걸쳐 동시적으로 실현된다. 이를 온전히 기록하기 위해서는 여러 대의 카메라를 다른 각도에서 동기화하여 촬영하는 것이 필수적이다. 그러나 이러한 기술적 요구는 필연적으로 녹화 환경의 인위성을 높여, '연구자의 존재나 녹음 장비가 화자의 언어 사용에 영향을 미칠 수 있다'는 고전적인 '관찰자 역설'의 문제를 음성언어 연구보다 훨씬 더 심화시킨다. 수어 연구자들은 이러한 딜레마를 해결하기 위해, 연구자가 없는 상황에서 참여자들끼리 직접 녹화하도록 하거나, 농인 연구자를 투입하여 참여자와의 심리적 거리를 좁히는 등의 방법론적 해법을 모색한다. 한국수어 말뭉치 보고서에 상세히 기술된, 최소 3대의 카메라와 조명으로 구성된 촬영실의 모습은 이러한 기술적 요구가 일상적인 대화 환경과는 다른 스튜디오 환경을 만들어냄을 명확히 보여준다(국립국어원, 2024: 45). 이는 음성언어 말뭉치 수집에 비해 수어 말뭉치 연구에서 관찰자 역설의 문제가 훨씬 더 근본적인 도전 과제가 됨을 의미한다.

두 번째 딜레마는 ‘윤리적 보호’와 ‘데이터의 자연성’ 사이의 긴장 문제이다. 민감정보 처리, 특히 참여자의 얼굴이 노출되는 영상 데이터의 윤리적 문제는 비단 수어 말뭉치만의 딜레마가 아니라, 관련 멀티모달 말뭉치 전반이 안고 있는 근본적인 제약 조건이다. 그러나 이 보편적인 딜레마는, 얼굴 정보 자체가 언어의 핵심적인 문법 요소로 기능하는 수어 말뭉치 연구에서 훨씬 더 복합적이고 첨예한 양상으로 드러난다.

농인 커뮤니티는 상대적으로 규모가 작고 긴밀하기에, 참여자의 익명성을 보장하는 것이 다른 언어 공동체에 비해 훨씬 더 중요하고 어려운 과제이다. 음성 데이터와 달리, 영상 데이터는 얼굴을 블러 처리하는 것만으로 문법적으로 중요한 정보(비수지 신호)가 심각하게 훼손될 수 있다. 이는 참여자를 윤리적으로 보호하려는 필수적인 ‘통제’가, 언어 현상의 ‘자연스러운’ 증거를 제거하는 결과를 낳을 수 있음을 의미한다. 이 딜레마에 대응하기 위해, 많은 해외 수어 말뭉치 프로젝트들은 일반 대중에게는 익명화된 일부 데이터만 공개하고, 엄격한 심사를 통과한 연구자에게만 제한적으로 원본 데이터에 접근 권한을 부여하는 계층적 접근(tiered access)모델을 해결책으로 채택하고 있다. 한국수어 말뭉치 보고서 역시 이러한 윤리적 고민을 반영하여, 법적·도덕적 문제나 특정인에 대한 부정적 내용이 포함된 영상을 ‘공개 불가’로 처리하는 등 엄격한 윤리적 기준을 적용하고 있음을 상세히 기술하고 있다(국립국어원, 2024: 102, 106-107).

5.6. 소결

본 장에서는 음성언어 말뭉치와는 다른 고유한 방법론적 과제를 안고 있는 수어 말뭉치를 ‘통제’와 ‘자연성’이라는 분석틀을 통해 심층적으로 고찰했다. 5.1절에서는 시각 채널의 특수성, 수어 문법의 동시성, 그리고 소리 정보 수집의 필요성 등 수어 말뭉치 구축의 근본적인 과제를 제시했다. 이어서 5.2절에서는 해외 주요 수어 말뭉치들의 사례를 통해, 자연스러운 대화와 통제된 유도 과제를 혼합하는 것이 일반적인 경향임을 확인했다. 5.4절과 5.5절의 분석을 통해, 이러한 경향이 두 패러다임 사이의 긴장을 해소하는 것이 아니라 오히려 더 복합적인 딜레마를 드러냄을 확인했다. 데이터의 정밀도를 높이기 위한 다중 카메라 시스템과 같은 기술적 ‘통제’는 필연적으로 관찰자 효과를 심화시켜 ‘자연성’을 위협한다. 또한, 농인 커뮤니티의 특수성을 고려한 ‘윤리적 보호’라는 또 다른 형태의 ‘통제’는, 데이터의 진정성을 일부 희생시킬 수 있다는 딜레마를 낳는다. 이는 수어 말뭉치 연구자들이 일반적인 방법론적 절충을 넘어, 기술적·

윤리적 딜레마 속에서 끊임없이 해법을 모색하고 있음을 보여준다.

결론적으로, 수어 말뭉치 분야는 본고에서 다루는 두 패러다임의 긴장이 가장 역동적이고 첨예하게 교차하는 영역이라 할 수 있다. 한국수어 말뭉치 역시 이러한 국제적 흐름과 고민을 공유하고 있음은 앞서 간략히 살펴보았다. 다음 장에서는 지금까지의 음성언어 및 수어 말뭉치 분석을 모두 종합하여, 향후 말뭉치를 설계하는 연구자들이 자신의 연구 목적에 맞춰 두 패러다임 사이의 긴장 관계를 어떻게 생산적으로 고려할 수 있는지에 대한 분석적 프레임워크를 제시하며 전체 논의를 마무리하고자 한다.

6. 결론

본고는 멀티모달 말뭉치의 발전사를 과학적 지식을 구성하는 두 가지 핵심 패러다임, 즉 내적 타당성을 추구하는 ‘통제된 데이터’와 외적 타당성을 지향하는 ‘자연성’ 사이의 인식론적 긴장 관계라는 렌즈를 통해 비판적으로 재해석했다. 4장과 5장의 분석을 통해, 음성언어 말뭉치와 수어 말뭉치 모두가 이러한 긴장 관계 속에서 각자의 연구 목적을 달성하기 위해 특정한 방법론적 선택을 해왔음을 확인했다. M2VTS와 같은 초기 말뭉치가 명확한 인과관계 규명을 위해 현실을 극도로 단순화하는 ‘통제’를 선택했다면, FOLK와 CEJC와 같은 후기 말뭉치들은 현실 세계의 복잡성을 있는 그대로 기술하기 위해 ‘자연성’을 추구했다.

이러한 분석은 ‘통제된 데이터’와 ‘자연성’이 어느 한쪽이 우월하거나 다른 한쪽으로 대체되는 선형적 발전 관계가 아님을 명확히 보여준다. 오히려 두 패러다임은 연구자가 어떤 종류의 ‘진실’을 추구하는가에 따라 선택해야 하는 상보적 관계에 놓여있다. 본고의 전체 논의를 바탕으로, 향후 멀티모달 말뭉치를 기획하는 연구자들이 자신의 연구 목적에 맞춰 고려할 수 있는 세 가지 설계의 가능성을 다음과 같은 분석적 프레임워크로서 제시하며 결론을 맺는다.

첫째, 연구 목적이 특정 변수 간의 명확한 인과관계 규명이라면, 연구자는 내적 타당성을 극대화하기 위해 변수를 엄격하게 통제하고 절차를 표준화하는 ‘통제된 데이터’ 패러다임을 따라야 한다. 둘째, 연구 목적이 특정 언어 공동체의 상호작용 양상을 있는 그대로 기술(description)하고 이해하는 것이라면, 대표성, 균형성, 진정성을 기준으로 삼는 ‘자연 데이터’ 패러다임에 기반하여 외적 타당성을 극대화해야 한다. 셋째, 연구 목적이 특정 이론을 실제 상황에 적용하

거나 복잡한 현실 문제 해결(예: AI 성능 향상)에 있다면, 두 패러다임의 절충이 필요하다. 이 경우, 연구자는 어떤 측면의 타당성을 어느 정도 수준에서 타협했는지를 명확히 인지하고 기술해야 한다.

궁극적으로, 이 프레임워크는 향후 한국어 멀티모달 말뭉치를 포함한 모든 말뭉치 기획자들이 자신의 연구 질문에 가장 적합한 방법론적 선택을 내리는 데 유용한 길잡이가 될 수 있을 것이다.

참고문헌

국립국어원 (2024). 「2024년 한국수어 말뭉치 수집 및 분석」. 서울 : 국립국어원.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd edition.). Berkeley, CA: University of California Press.

Busso, C., Bulut, M., Lee, C. C., Kazemzadeh, A., Mower, E., Kim, S., Chang, J., Lee, S., & Narayanan, S. S. (2008). IEMOCAP: Interactive emotional dyadic motion capture database. *Language Resources and Evaluation, 42*(4), 335-359.

Carletta, J., Ashby, S., Bourban, S., Flynn, M., Guillemot, M., Hain, T., ... & Wellner, P. (2005). The AMI meeting corpus: A pre-announcement. In *International Workshop on Machine Learning for Multimodal Interaction,* (pp. 28-39). Berlin: Springer.

Clayman, S. & Heritage, J. (2002). *The news interview: Journalists and public figures on the air.* Cambridge: Cambridge University Press.

Crasborn, O. & Zwitserlood, I. (2008). The Corpus NGT: An online corpus for professionals and laymen. In *Proceedings of the LREC 2008 Workshop on the Representation and Processing of Sign Languages,* (pp. 38-43).

Geertz, C. (1973). *The interpretation of cultures.* New York, NY: Basic Books.

Goodwin, M. H. (1990). *He-said-she-said: Talk as social organization among black children.* Bloomington, IN: Indiana University Press.

Jahn, E., Konrad, R., Langer, G., Wagner, S., & Hanke, T. (2018). Publishing DGS corpus data: Different formats for different needs. In M. Bono, E. Efthimiou, S.-E. Fotinea, T. Hanke, J. A. Hochgesang, J. Kristoffersen, J.

Mesch, & Y. Osugi (Eds.), *Proceedings of the LREC2018 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community,* (pp. 83-90). European Language Resources Association (ELRA).

Jefferson, G. (1984). Transcription notation. In J. Maxwell Atkinson & John Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. ix-xvi). Cambridge: Cambridge University Press.

Johnston, T. (2008). Auslan Corpus. Endangered Languages Archive. http://hdl.handle.net/2196/00-0000-0000-0000-D7CF-8

Johnston, T. (2010). From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics, 15*(1), 104-129.

Koiso, H., Amatani, H., Den, Y., Iseki, Y., Ishimoto, Y., Kashino, W., Kawabata, Y., Nishikawa, K., Tanaka, Y., & Usuda, Y. (2022). Design and evaluation of the corpus of everyday Japanese conversation. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Thirteenth Language Resources and Evaluation Conference,* (pp. 5587-5594). European Language Resources Association (ELRA).

Koiso, H., Den, Y., Iseki, Y., Kashino, W., Kawabata, Y., Nishikawa, K., Tanaka, Y., & Usuda, Y. (2018). Construction of the corpus of everyday Japanese conversation: An interim report. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis, & T. Tokunaga (Eds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018),* (pp. 469-476). European Language Resources Association (ELRA).

Labov, W. (1972). *Sociolinguistic patterns.* Philadelphia: University of Pennsylvania Press.

Messer, K., Matas, J., Kittler, J., Luettin, J., & Maitre, G. (1999). XM2VTSDB: The extended M2VTS database. In *Second International Conference on Audio and Video-based Biometric Person Authentication (AVBPA'99).*

Neidle, C. (2007). SignStream: A database tool for research on visual-gestural language. *Sign Language and Linguistics, 10*(2), 203-214.

Peräkylä, A. (1995). *AIDS counselling: Institutional interaction and clinical practice.* Cambridge: Cambridge University Press.

Pigeon, S. & Vandendorpe, L. (1997). The M2VTS multimodal face database. In Josef Bigün, Gérard Chollet, & Gunilla Borgefors (Eds.), *Audio- and video-based biometric person authentication,* (pp. 403-409). Berlin: Springer.

Prillwitz, S., Hanke, T., König, S., Konrad, R., Langer, G., & Schwarz, A. (2008). DGS corpus project—Development of a corpus based electronic dictionary German Sign Language/German. In O. Crasborn, E. Efthimiou, T. Hanke, E. D. Thoutenhoofd, & I. Zwitserlood (Eds.), *Proceedings of the LREC2008 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora,* (pp. 159-164). European Language Resources Association (ELRA).

Reece, A., Cooney, G., Bull, P., Chung, C., Dawson, B., Fitzpatrick, C., Glazer, T., Knox, D., Liebscher, A., & Marin, S. (2023). The CANDOR corpus: Insights from a large multimodal dataset of naturalistic conversation. *Science Advances, 9*(13), eadf3197.

Sacks, H. (1992). *Lectures on conversation* (G. Jefferson, Ed.). Oxford: Blackwell.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4), 696-735.

Schmidt, T. (2014). The research and teaching corpus of spoken German FOLK. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14),* (pp. 383-387). European Language Resources Association (ELRA).

Schmidt, T. (2016). Good practices in the compilation of FOLK, the research and teaching corpus of spoken German. *International Journal of Corpus Linguistics, 21*(3), 396-418.

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: Oxford University Press.

ten Have, P. (2007). *Doing conversation analysis: A practical guide* (2nd edition.). London: Sage Publications.

Wang, Z., Wu, S., Chen, H., He, M. K., Du, J., Lee, C. H., ... & Liu, C. (2023). The multimodal information based speech processing (MISP) 2022 challenge: Audio-visual diarization and recognition. *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*.

조용준
05029 서울시 광진구 능동로 120
건국대학교 다언어다문화연구소 학술연구교수
이메일: yjcho33@naver.com

안희돈
05029 서울시 광진구 능동로 120
건국대학교 영어영문학과 교수
전 화: 02-450-3338
이메일: hdahn@konkuk.ac.kr

논문접수: 2025년 11월 18일
수정완료: 2025년 12월 26일
게재확정: 2025년 12월 31일