

텍스트 마이닝을 활용한 국내 제약·바이오 업종의 EU 공급망 실사법 대응 방안 연구

A Study on the Countmeasures of the Korean Pharmaceutical/ Bio Industry to the EU Corporate Sustainability Due Diligence Directive, by using Text Mining

김 소 리 (Sori Kim) 한국표준협회 ESG경영센터 연구원, 교신저자
기 준 학 (Joonhak Ki) 숙명여자대학교 산학협력단 산학협력교수

요 약

EU는 2022년 2월, 기업 공급망의 환경·인권 리스크 실사 이행 및 정보공시를 의무화하는 EU 기업 지속가능성 실사 지침(EU Corporation Sustainability Due Diligence Directive) 초안을 발표하였다. 추후 해당 지침의 최종안이 확정 및 발효되면 EU 회원국은 2년 내에 이를 자국법으로 제정하여야 한다.

본 연구에서는 EU가 요구하는 공급망 실사 의무 요건에 대한 13개 국내 제약·바이오 기업의 대응 현황을 살펴보고, 환경·인권 영향 관리를 잘하고 있다고 평가받는 13개 글로벌 선진 기업의 대응 현황과 비교 분석하였다.

비교 분석을 위해 R을 활용해 텍스트 마이닝 분석을 수행하였다. 기본적인 단어빈도와 동시출현단어를 분석하였고 Latent Dirichlet Allocation 기법을 적용해 토픽 모델링을 수행하였다.

분석 결과, 선진 기업 대비 국내 제약·바이오 기업은 부정적 이슈의 식별 및 보고 체계와 공급망 실사 이행 프로세스가 미흡하며, 환경·인권 정보공시를 위한 데이터 관리 고도화가 필요한 것으로 나타났다. 또한 국내 제약·바이오 기업은 단순히 자금 지원 등의 수준을 넘어 중소기업 공급망이 가진 리스크를 체계적으로 파악하고, 이를 저감할 수 있는 차별화된 지원방안을 마련하여야 할 것으로 판단된다. 정부 역시 전문가 컨설팅, 재원 지원 등 국내 제약·바이오 기업 공급망 실사 대응 역량 강화를 위한 지원과 함께 우리나라의 독자적인 공급망 환경·인권 관리제도 의무화 시행을 통해 정책적인 뒷받침을 강화할 필요가 있다.

키워드 : ESG경영, 텍스트마이닝, EU 공급망 실사법, 제약바이오 업종, 선진기업 비교분석

† 이 논문은 김소리의 2023년도 석사학위논문 ‘텍스트마이닝을 활용한 국내 기업의 EU 공급망 실사법 대응 방안 연구-철강 및 제약·바이오 업종을 중심으로’의 일부를 발췌하여 재작성한 것으로 환경부 재원의 ‘환경산업기술원 지식기반 특성화대학원 사업’ 지원을 받아 수행된 연구임

I. 서 론

유럽연합(European Union, 이하 EU)은 무역 의 제 우선순위를 지속가능성 관련 글로벌 행동 중심 으로 설정하고(Okcu, 2022) 다양한 환경 · 인권 관 련 무역 규제를 발표하며 규칙 형성자(rule-maker) 로서의 역할을 강화하고 있다. 특히 ESG와 무역 이슈를 연계해 법률 및 정책적으로 담론화하고 있 다(Kim and Choi, 2022).

2022년 2월, EU는 일정 규모 이상의 기업을 대상으 로 자사 및 공급망 내 환경 · 인권 리스크의 확인 · 보 고 · 개선 의무를 법제화하는 EU 기업 지속가능성 실사 지침(Corporate Sustainability Due Diligence Directive, 이하 CS3D)의 초안을 공개하였다.

CS3D는 기업의 가치사슬 내에 있는 전 협력사의 환경 · 인권 관련 부정적 영향을 파악해야 하는 만 큼 대응 범위가 넓고, 대기업은 물론 중소기업까지 해당 지침 적용의 직간접 영향에 들 수 있어 파급력 이 큰 제도이다. 황준석, 장현숙(2023)은 파급력 ·

시급성 · 대응난이도를 종합적으로 고려했을 때 국 내 기업 영향이 가장 큰 환경규제가 CS3D(‘공급망 실사 지침’)인 것으로 지목하였다(<표 1>).

이러한 규제는 우리 기업에 기회가 될 수도 있 다. 이수열, 이준겸(2015)은 공급사가 글로벌 원청 기업의 환경 · 사회적 요구 수준을 긍정적 관점으 로 받아들일 필요가 있음을 제시하였다. 공급사가 자국의 쉬운 환경 · 사회 규제 및 기준만을 충족하 는 것은 글로벌 경쟁의 측면에서 의미가 없으며, 글로벌하게는 가장 엄격한 환경 · 사회적 기준이 통용될 가능성이 높으므로 이에 대한 대응 역량을 높임으로써 공급 범위를 확대하고 경제적 성과 개 선을 이룰 수 있기 때문이다.

강노경(2022) 역시 EU에서 CS3D를 시행하게 되면 실사 요구를 준수할 수 있는 기업을 중심으 로 공급망이 재편될 것으로 전망하며, 이에 대응 할 수 있는 경영활동 및 정보공개시스템을 구축해 현재 중국에 편중된 EU 공급망에 우리 기업이 진출하는 기회로 활용해야 한다고 주장하였다.

<표 1> EU 환경규제별 영향 비교

주요 규제	개요		국내 기업 영향			
	발효시기	적용대상	종합	파급력	시급성	대응 난이도
탄소국경조정제도 (CBAM)	2023 상반기	철강, 알루미늄, 비료, 시멘트, 전력, 수소	●	●	○	●
공급망실사 지침	2023 하반기	4천만 유로 초과 매출기업 (고위험산업군 비중 50% 이상)	●	●	○	●
지속가능성 공시 지침 (CSRD)	2023.1.	1.5억 유로 초과 매출기업 /4천만 유로 초과 매출지사 또는 동 지침 적용 자회사가 있는 기업	●	●	●	○
에코디자인 규정 (ESPR)	2023 중	모든 물리적 상품	●	●	○	●
新배터리 규정	2023 상반기	이동식배터리, 전기차·전기자전거 배터리, 2kWh 이상 충전식 산업용 배터리	○	○	○	○
플라스틱세	2023.1.	플라스틱	●	○	●	○
화학물질 분류 및 포장 규정(CLP)	2023 상반기	화학	○	○	●	○
EU REACH&RoHS	2023.1.	전기전자제품 및 화학	○	○	●	○

주: 파급력, 시급성, 대응 난이도를 기준으로 정성평가
 자료: 연구자 분석

● 매우 높음 ● 높음 ○ 보통 ○ 낮음 ○ 매우 낮음

II. 연구방법

2.1 연구목적

본 연구는 공급망 ESG 관리 요구가 강화되는 글로벌 추세 속에서 국내 기업이 CS3D라는 새로운 ESG 규제에 얼마만큼의 대응력을 갖추고 있는지 현황을 파악하고 그에 따른 대응 방안을 제시하고자 한다. 이를 통해 환경·인권 등 ESG 역량에 따라 재편될 국제 무역 질서에서 우리 기업의 경쟁력 훼손을 막는 데 기여할 수 있을 것으로 기대한다.

이에 기업 ESG 경영 체계와 성과를 망라한 지속가능경영보고서(sustainability report)를 텍스트 마이닝으로 분석해 동일 업종의 국내 기업과 글로벌 ESG 선진기업의 ESG 경영 현황을 비교함으로써 CS3D의 요구사항 대비 국내 기업의 개선방안을 도출하고 공급망 실사 대응을 위한 정책적 시사점을 제시하고자 한다.

또한, ESG 및 공급망 실사 대응 역량을 파악하기 위해서는 각 기업의 실사 프로세스, 이행 조직 등 수치화할 수 없는 정성적 요소에 대한 평가도 요구된다. 이에 비정형 데이터인 텍스트를 정량적으로 분석할 수 있는 텍스트 마이닝이 새로운 규제에 대한 기업의 대응 현황을 파악하고 개선방안을 도출하는 데 유용한 연구방법론임을 입증하고자 한다.

2.2 연구대상

제약·바이오 업종은 한국생산성본부가 꼽은 CS3D 최우선 영향 업종 중 하나로(박병립, 2022), 세계 2위 시장인 유럽으로의 제품 및 기술 수출이 중요하게 작용하는 업종이다. 그러나 제약·바이오 업종의 수출액은 최근 지속적으로 감소하고 있으며(산업통상자원부, 2023), 한국ESG기준원(2022)의 평가 결과에서도 ESG 대응 역량이 취약한 업종인 것으로 나타났다. 2022년 한국ESG기준원의 상

장사 ESG 평가에서 90개의 제약·바이오 업종 기업 중 ESG 통합 등급 우수기업은 5개사에 불과한 것으로 나타났으며 ESG 등급 취약기업은 28개사, 매우 취약기업은 28개사로 확인되었다.

이에 본 연구에서는 제약·바이오 업종을 연구 대상으로 선정하고, 국내 제약·바이오 업종 중 의약품 물질 및 의약품 제조 기업을 대상으로 2021년 매출액 기준 상위 50개사를 선별, 그중에서 지속가능경영보고서를 발간하고 있는 13개 기업을 대상으로 연구를 진행하였다(<표 2>).

글로벌 선진 제약·바이오 기업은 2021년 매출액 기준 글로벌 상위 20개 기업 중에서 모건 스탠리 캐피탈 인터내셔널(Morgan Stanley Capital International) A 등급 이상(MSCI, 2023), 탄소정보공개프로젝트(Carbon Disclosure Project) 항목별 A 등급 이상(CDP, 2023), 코퍼레이트 나이즈(Corporate Knights) 선정 글로벌 지속가능성 100대 우수기업(Corporate Knights, 2023) 포함 등 세 가지 기준 중 하나 이상을 충족하는 기업 중 상위 13개 기업으로 선정하였다(<표 2>).

<표 2> 제약·바이오 업종 연구대상

	국내 제약·바이오 기업	글로벌 선진 제약·바이오 기업
1	유한양행	Pfizer
2	삼성바이오로직스	Johnson&Johnson
3	녹십자	Novartis
4	대웅제약	Roche
5	씨젠	Merck&Co.
6	한미약품	Bristol-Myers Squibb
7	에스케이바이오사이언스	Sanofi
8	에이치케이이노엔	AstraZeneca
9	보령	Takeda
10	서흥	Gilead
11	동아에스티	Amgen
12	휴온스글로벌	Bayer
13	제뉴원사이언스	Eisai

2.3 분석 데이터

본 연구는 지속가능경영보고서를 기초 데이터로 활용한다. 지속가능경영보고서는 지속 가능한 개발 목표를 향한 조직의 성과를 측정·공개·설명하는 수단으로 조직이 경영 활동 전 과정에서 추진한 주요 노력과 성과를 공개하는 보고서이다(한국품질재단, 2022). 이해관계자는 지속가능경영보고서를 통해 해당 기업의 지속가능경영 방향성과 체계, 구체적인 환경·사회·지배구조 개선 노력을 파악할 수 있다.

지속가능경영보고서는 기업이 임의 기준대로 만드는 것이 아니라 하나 또는 여러 개의 글로벌 비재무정보 공시 표준 및 가이드라인을 기반으로 작성된다. 대표적인 공시 기준으로는 Global Reporting Initiative(이하 GRI)를 비롯해 2018년 제정된 미국 지속가능성 회계기준위원회(Sustainability Accounting Standards Board) 표준, 기후 관련 재무정보 공시 태스크포스(Task Force on Climate-related Financial Disclosure) 등이 있다.

지속가능경영보고서를 분석 데이터로 삼은 이유는 CS3D의 요구사항인 기업의 인권 및 환경 관련 부정적 이슈 관리가 결국 환경, 인권, 지배구조를 다루는 ESG 정보와 유사하기 때문이다. 또한, 앞서 기술했듯이 지속가능경영보고서는 글로벌 공시 표준과 가이드라인을 기반으로 해당 기업의 ESG 추진 목적, 체계, 활동, 성과 등을 총망라해 적게는 수십 페이지에서 많게는 수백 페이지 분량

의 정보를 담고 있으므로 기업 간 ESG 경영 현황을 동일한 표준에 따라 일관성 있게 비교·분석하는 기준 데이터가 될 수 있기 때문이다.

본 연구의 목적상 ESG 추진 내용의 시계열적 변화 과정이 아닌 현시점에서의 업종별 기업 ESG 추진 현황을 비교 분석하는 것이 중요하므로 2023년 1분기 기준, 대상 기업의 최신 발간 보고서를 분석 대상으로 선정하였다. 국내 기업은 국문 지속가능경영보고서를, 글로벌 선진 기업은 영문 지속가능경영보고서를 활용하였다. 일관성 있는 키워드 도출을 위해 통합보고서는 제외하고 지속가능경영보고서만을 분석 대상으로 삼았다. 또한 데이터 간 비교 적합성을 높이기 위해 ESG와 관련한 별도의 기업 내부 규정이나 지침, 규범, 성명서, 홈페이지에 게재된 ESG 경영 관련 텍스트 등은 제외하고 지속가능경영보고서만을 대상으로 분석을 수행하였다.

2.4 분석방법

텍스트 마이닝은 비정형 텍스트 데이터에 자연어 처리 기술을 적용해 대량의 텍스트 데이터에서 의미 있는 정보를 추출, 가공하는 데이터 분석 기법이다(김성근 등, 2016; 전효정 등, 2023). 연구자가 직접 텍스트를 읽고 분석하면 주관이 반영되지만 텍스트 마이닝을 이용하면 주관을 배제하고 편견 없이 텍스트의 내용을 파악할 수 있다. 텍스트 마이닝에는 <표 3>에 나타난 바와 같이 다양한 분석 기법이 활용된다(김영우, 2021).

<표 3> 텍스트 마이닝의 다양한 분석 기법

분석 기법	내용
단어 빈도 분석 (Keyword Frequency Analysis)	단어가 문서에 몇 번 사용되었는지 알아보는 분석 방법으로 자주 사용된 단어를 통해 해당 문서가 어떤 내용을 강조하는지 확인
의미 연결망 분석 (Semantic Network Analysis)	연결된 단어의 관계를 통해 문서를 해석하는 분석 방법으로 동시출현단어 분석(Co-occurrence Analysis), 중심성 분석(Centrality Analysis) 등이 있음
감정 분석 (Sentiment Analysis)	단어 긍·부정 사전을 이용해 텍스트에 담긴 감정을 분석하는 방법으로 해당 문서가 어떤 감정으로 분류되는지 분석
토픽 모델링 (Topic Modeling)	텍스트의 핵심 주제를 찾아 비슷한 내용끼리 분류함으로써 텍스트의 주요 주제 군집 도출

텍스트 마이닝은 ESG 관련 연구에서 특히 유용하게 활용할 수 있는 분석방법론이다. 기업 비재무정보의 경우 수치화된 정량 데이터뿐 아니라 경영 체계, 운영 조직, ESG 각 분야별 체계와 활동 등 정성 데이터의 비중이 높기 때문이다.

본 연구에서는 통계분석프로그램 R과 Rstudio를 활용해 텍스트 마이닝 분석을 수행하였다. 한글 텍스트(n≥2) 분석을 위해 rJAVA 프로그램과 KoNLP 의존성 패키지를 활용하였고, 영어 텍스트 분석은 snowballC 패키지를 사용해 분석하였다. 더 이상 분리되지 않는 의미의 최소 단위인 형태소를 기준으로 토큰화하였으며, 영어 텍스트의 경우 같은 의미의 단어라도 품사나 시제에 따라 형태소가 달라질 수 있기 때문에 이를 모두 다른 단어로 인식하지 않도록 의미 단위인 어간(stem)을 기준으로 토큰화해 분석하였다. 의미가 비슷한 일부 단어들, 예컨대 ‘공급망’·‘공급사’·‘협력사’와 ‘기업’·‘회사’는 각각 ‘공급망’, ‘기업’ 같은 대표 단어로 통합하여 분석하였다.

III. 이론적 배경 및 선행연구

3.1 EU 기업 지속가능성 실사 지침 주요 내용

CS3D는 기존에 유럽 내 비재무정보 공시지침과

더불어 「OECD 다국적 기업 가이드라인(OECD, 2018)」, 「UN 이행 원칙 보고 프레임워크(UNGP, 2017)」 등 국제적으로 합의된 자발적 ESG 공시 및 실사 표준의 한계를 보완하기 위해 제정되었다.

자발적 공급망 실사에는 실사 의무의 불명확성, 가치사슬의 복잡성, 법적 명확성의 부족 등 여러 한계가 내재되어 있다. 또한 EU 회원국별 공급망 실사 의무화 법안 제정 시 회원국 간 상이한 요구사항으로 인해 기업 가치 상충이나 혼란, 비용 발생 등을 초래해 불공정 경쟁이 발생할 위험이 있다. 이에 대응하기 위해 EU 차원의 CS3D가 수립되었다(European Commission, 2022).

EU는 CS3D 의무화를 통해 기업이 복잡한 공급망 내의 환경·인권 위험을 식별하고 제거·완화하여 지속 가능한 경영활동을 추진할 수 있는 기반을 제공한다. CS3D에서 규정하는 주요 내용은 <표 4>와 같다(남승현, 2022).

또한 <표 5>에 CS3D에 따라 기업이 수행하여야 하는 기업의 실사 활동을 단계별로 나누어 나타내었다.

기업이 공급망 실사 의무를 소홀히 하거나, 실사를 이행하였음에도 공급망 내에서 중대한 위험이 발생하고 기업이 이를 식별·완화하는 노력을 적절히 수행하지 못한 경우에 기업은 벌금 등 행정 제재 또는 민사상 배상 청구 대상이 될 수 있다.

<표 4> CS3D의 주요 내용

적용 대상 기업 (지침 제1조 및 제3조)	- 자회사를 포함한 전사 경영 활동 및 당해 기업 가치 사슬(value chain)에 포함된 업체 중 확립된 비즈니스 관계(established business relationship)를 중심으로 실사 수행 - 매년 자사 가치사슬 내 어떤 협력 업체를 확립된 비즈니스 관계로 볼 것인지 검토 필요
기업의 의무 (지침 제4조 ~제11조)	- 전 공급망 활동에서 발생 가능한 인권·환경 관련 부정적 영향을 식별하고 예방 및 완화, 대응 - 기업은 인권 및 환경 실사를 위한 정책을 수립 및 내재화하고, 이에 따라 공급망 기업의 실질적·잠재적·부정적 영향을 식별하며, 식별된 영향에 대한 방지·완화·개선 또는 종료·최소화 실행계획을 수립 및 이행 - 제3자 검증 및 산업 이니셔티브를 활용해 협력사의 실사 계약 준수 여부를 지속적으로 확인, 부정적 영향을 예방하거나 완화할 수 없는 경우 해당 협력사와 사업 관계를 중단하거나 새로운 관계를 맺도록 명시 - 이해관계자가 기업에 적절한 후속 조치를 요청할 수 있도록 고충처리제도 마련, 매년 1회 이상 정기적으로 모니터링을 실시해 실사 정책의 효율성과 적절성 점검 - 일련의 실사 정책 및 조치 시행 후 매년 정해진 기한까지 실사 내용 공시

〈표 5〉 공급망 실사 단계

1	2	3	4	5	6	7
실사 정책 내재화	영향 식별	잠재적 영향 예방·완화	실질적 영향 종료·최소화	고충처리 제도화	모니터링	정보공시

CS3D 지침은 EU에서 활동하며 고용인원 및 매출 기준을 충족하는 역내·외 대기업과 중견기업에 적용된다. 도입 첫해부터 임직원 250명, 매출 4,000만 유로 초과 기업에 해당 지침을 적용하도록 명시하고 있다. 또한 임직원 500명 초과, 전 세계 매출 1억 5,000만 유로 초과 모기업에도 해당 규제가 적용된다. EU 역외 기업의 경우 EU 내 매출 4,000만 유로 초과, 글로벌 매출 1억 5,000만 유로 이상이면 적용된다.

EU 역내 기업 1만 3,800개 사, 역외 4,000개 사에 해당 지침이 적용될 것으로 예상되며, 한국의 수출 기업도 직간접적으로 광범위하게 영향을 받을 것으로 예상된다(남승현, 2022). 국내 기업의 CS3D 적용 대상 유형은 <표 6>과 같이 나눌 수 있다(관계부처 합동, 2021).

앞서 언급한 기준을 충족하지 않는 중소기업의 경우 지침의 직접 적용 대상은 아니지만, 유럽 기업과 거래하거나 유럽에 제품을 수출하는 경우 원청사의 실사 준수 계약 요구에 따라 간접적인 영향을 받을 수 있다. 때문에 중소기업 역시 인권·환경 위험을 예방하고 ESG 경영 역량을 제고하는 등 공급망 실사에 대응하는 노력이 필요하다.

CS3D 최종안 선택 및 지침 발효 시점은 2024년 중으로 예상되며, 지침이 통과되면 2년 내에 각 회원국 입법 및 규제 적용이 이루어질 예정이다.

EU 시장에 진출해 있는 국내 대기업은 물론 중견·중소 협력사 역시 이에 대응하기 위해서는 인권·환경 경영 수준에 대한 진단 및 CS3D 지침이 요구하는 구체적인 사항에 대한 이행 역량을 갖추어야 한다.

3.2 국내 기업의 공급망 실사 대응 현황

2023년 2월, 대한상공회의소(2023)에서 국내 300개 기업을 대상으로 2023년 ESG 주요 현안과 정책 과제를 조사한 결과 ‘공급망 실사 대응’이 40.3%로 ‘ESG 의무공시(30.3%)’, ‘순환경제 구축(15.7%)’, ‘탄소국경조정제도(12%)’를 누르고 현안 과제 1순위로 꼽혔다(<그림 1>). 그러나 실사 대응 수준을 묻는 항목에는 절반 가까운 기업이 ‘별도 조치 없음’으로 응답해(원청사 48.2%, 협력사 47%) ESG 공급망 실사와 관련한 지원 대책 마련이 필요한 상황임을 알 수 있다(<표 7>).

이에 산업통상자원부와 한국생산성본부는 2021년 12월 기업의 공급망 실사 자가진단을 지원하기 위해 「공급망 실사 대응을 위한 K-ESG 가이드라인」을 발간하였다. 2023년 2월에는 ESG 민관합동 컨트론타워인 ESG 협의회에서 정부가 수출 중소·중견기업의 ESG 공급망 실사 대응 역량 확보를 위한 진단평가 및 컨설팅 사업을 추진하겠다고 밝혔다.

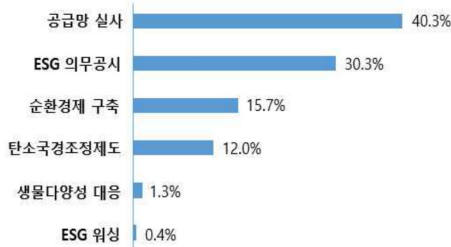
〈표 6〉 EU의 기업 지속가능성 실사 지침 적용 대상 중 국내 기업 해당 유형

구분	직접 적용 기업 : 공급망 ESG 직접 실사 후 결과 보고 의무 기업	간접 적용 기업 : 공급망 실사 결과보고 의무기업의 국내 협력사
대상	- EU 현지 법인 보유 혹은 EU 수출 규모가 일정 수준 이상인 기업 - 일정 규모 이상 고용 및 매출 기준 적용	- 실사 결과 보고 의무를 가진 현지 기업과 거래하는 기업 - EU 및 비EU 기업 대상 수출 기업 중 공급망 ESG 실사 잠재 대상
영향 시기	- 법안 발효 시 직접 영향(EU) 2024~2025년 예상(독일) 2023년 1월	- 각국 법안 발효 후 원사업자 판단에 따라 영향 시기 변동

〈표 7〉 ESG 실사법 대응현황 조사 결과

EU발 공급망 ESG 실사법에 어떻게 대응하십니까?

구분	[실사 이전 단계] 실사 시스템 구축	[본격 실사 단계] 온라인/현장 실사	[실사 후 피드백 단계] 실사 후 개선조치 요구	별도 조치 없음
원청기업	24.9%	14.5%	12.4%	48.2%
구분	대응 조직 구성	진단/평가 시행	진단/평가 후 컨설팅 시행	별도 조치 없음
협력업체	13.7%	26.2%	13.1%	47.0%



〈그림 1〉 2023 ESG 현안 응답

3.3 선행연구

ESG 경영 분석에 텍스트 마이닝을 적용하기 위한 기본 전제는 경영 전략, 중점 방향, 매출 수준 등 기업의 특성에 따라 지속가능경영보고서에 사용되는 단어가 다르게 출현한다는 것이다. 이를 근거로 기업의 주요 경영 전략 문서를 텍스트 마이닝으로 비교 분석하여 유의미한 시사점을 도출하는 연구들이 진행되고 있다.

Clatworthy(2006)는 매출에 따른 기업의 최고경영자의 메시지 차이를 분석한 결과, 수익성이 낮은 기업은 주요 재무지표 및 정량적 성과를 덜 언급하고, 수동태 문장을 덜 사용하며, 미래에 더 집중하는 경향을 보이는 등 사용하는 단어에 차이가 발생함을 확인하였다. Tanaka and Tsuda(2022)는 일본의 우수 통합보고서 수상 기업과 미수상 기업의 보고서 텍스트를 비교 분석하여 ‘기업가치’, ‘사업성장’, ‘비즈니스 플랜’ 등의 단어가 많이 출현하는 우수 보고서의 특징을 비롯해 두 그룹 간 차이점을 규명한 바 있다.

특히 업종에 따라 출현 단어의 특성이 달라지므로 벤치마킹을 위한 텍스트 비교 분석 시 연구 대상 업종을 구분하는 것이 필요하다. 박수빈 등(2022)은 은행, 보험, 증권 각 업종에서 ESG 역량이 우수한 2개 기업씩을 선정해 지속가능경영보고서를 텍스트 마이닝으로 분석한 결과, 동일한 금융업이라도 세부 업종별로 고빈도 출현 단어와 네트워크 그래프의 중심성이 다르게 나타남을 확인하였다. 김원희, 권영옥(2023)은 금융, 제조, IT 산업의 국내 주요 기업 ESG 보고서와 보도 기사를 대상으로 키워드 빈도 분석과 토픽 모델링을 수행하였으며, 그 결과 금융 산업은 ‘고객 중심 경영’ 및 ‘기후 변화 대응’, 제조 산업은 ‘지속 가능한 공급망 관리’ 및 ‘탄소중립’, IT 산업은 ‘기술 혁신’ 및 ‘디지털 책임’을 강조하는 등 업종별로 차이가 있음을 확인하였다. Sharma *et al.*(2022)은 텍스트 마이닝을 활용하여 Fortune 500대 기업의 사회책임보고서를 분석한 결과 기업의 국적, 규모, 업종, 운영 기간 등 다양한 요인에 따라 사회책임 보고의 초점이 상이함을 확인하였다. Abadi *et al.*(2018)은 Fortune 500대 기업 내 글로벌 식품 기업 및 이란 식품 기업의 사명과 비전 선언문(Mission and Vision Statement)을 텍스트 마이닝과 K-Means 알고리즘 클러스터링 등을 이용하여 분석함으로써 경영 전략의 차이점을 도출하였다. 글로벌 식품 기업은 ‘고객 창출’, ‘이미지 형성’에 중점을 두는 반면, 이란 식품 기업은 ‘제품’, ‘개발’, ‘일반적인 개념 서술’에 초점을 맞추고 있는 것으로 나타났다.

한가록 등(2022)은 텍스트 마이닝 방법을 활용하

여 EU의 CS3D 핵심 키워드를 추출하였다. 또한 ‘중소기업’, ‘공급망’ 키워드 관련 언론 기사를 분석해 국내 중소기업이 CS3D에 대응할 수 있는 방안을 제안하였다. 그러나 CS3D의 구체적인 요구사항에 대한 국내 기업의 대응 현황이나 업종별 분석 등은 해당 연구에서 고려되지 않았다.

본 논문에서는, 유럽 시장으로의 제품이나 기술 수출 중요성이 높은 제약·바이오 산업을 연구 대상 업종으로 선정하고, 한가록 등(2022)의 선행 연구에서 도출된 CS3D 핵심 키워드 대비 연구 대상 그룹의 CS3D 대응 현황을 평가하고자 한다. CS3D의 환경·인권 경영 요구사항을 비교적 잘 수행한다고 평가받는 글로벌 선진 기업과 국내 기업의 지속가능경영보고서를 다양한 텍스트 마이닝 기법을 활용하여 비교 분석하고, 두 그룹 간 분석 결과의 차이에서 시사점을 도출하는 벤치마킹 방식을 통해 국내 기업에 필요한 CS3D 대응 방안을 제시할 수 있을 것으로 기대한다.

IV. 분석결과

4.1 단어 빈도 분석

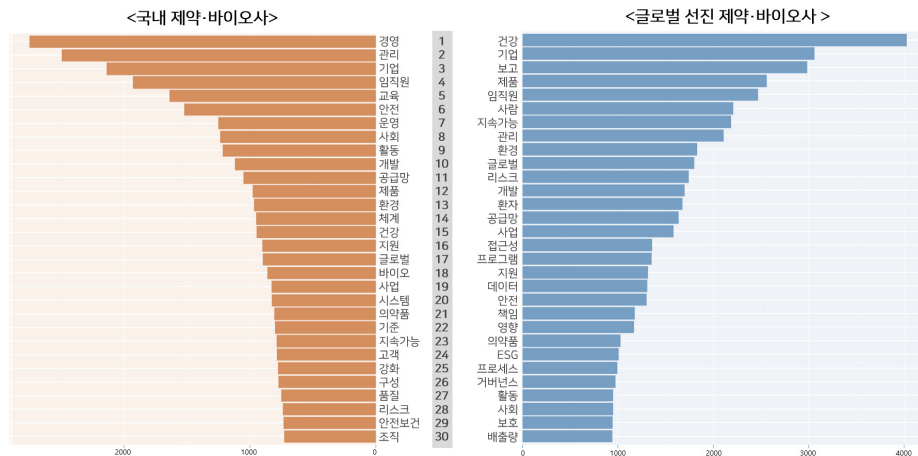
국내 제약·바이오사와 글로벌 선진 제약·바

이오사의 빈출 단어를 기반으로 구현한 워드 클라우드(word cloud)는 <그림 2>와 같다. 워드 클라우드는 단어빈도를 구름과 같은 형태로 시각화한 그래프로, 빈도에 따라 글자의 크기와 색을 다르게 표현해 어떤 단어가 얼마나 많이 사용되었는지 한눈에 파악할 수 있다(김영우, 2021).

<그림 2>의 워드 클라우드를 통해 국내 제약·바이오사에서는 ‘경영’, ‘관리’, ‘기업’, ‘임직원’ 등을 많이 사용하는 반면 선진 제약·바이오사의 경우에는 ‘건강’, ‘운영’, ‘보고’, ‘제품’ 등의 단어를 많이 사용한다는 것을 직관적으로 확인할 수 있다.



<그림 2> 국내 및 글로벌 선진 제약·바이오사 단어빈도 워드 클라우드



<그림 3> 제약·바이오 업종 단어빈도 그래프

두 연구 대상 그룹의 출현 단어별 빈도 순위는 <그림 3>과 같다. 비교의 용이성을 위해 글로벌 선진 제약·바이오사 분석 결과는 국문으로 번역하여 기술하였다.

<그림 3>에서 30위 내 고빈도 단어들의 출현횟수를 살펴보면 글로벌 선진 기업의 경우 900~4,100 사이로, 700~3,000 사이인 국내 기업보다 단어의 절대 출현횟수가 더 많은 것을 알 수 있다. 두 연구

대상 그룹 간 기업 수가 동일한 것을 고려할 때 글로벌 선진 기업이 더 많은 ESG 이슈 내용을 보고하고 있음을 보여준다.

제약·바이오 업종 단어 빈도 분석에 따른 구체적인 출현 빈도는 <표 8>과 같다.

두 그룹 모두 ‘기업’, ‘임직원’, ‘관리’ 등의 단어 빈도가 높게 나타났으며 ‘제품’, ‘개발’, ‘사업’ 등 비즈니스 관련 단어도 양쪽 모두 고르게 확인되었

<표 8> 제약·바이오 업종 단어 빈도 분석

출현빈도	국내 제약·바이오사 출현 단어	순위	글로벌 선진 제약·바이오사 출현 단어	출현빈도
2752	경영	1	건강	4027
2497	관리	2	기업	3060
2140	기업	3	보고	2983
1930	임직원	4	제품	2559
1639	교육	5	임직원	2468
1522	안전	6	사람	2209
1251	운영	7	지속가능	2186
1237	사회	8	관리	2109
1217	활동	9	환경	1831
1119	개발	10	글로벌	1801
1051	공급망	11	리스크	1742
979	제품	12	개발	1699
967	환경	13	환자	1677
952	체계	14	공급망	1637
947	건강	15	사업	1584
902	지원	16	접근성	1358
898	글로벌	17	프로그램	1354
861	바이오	18	지원	1315
829	사업	19	데이터	1307
826	시스템	20	안전	1301
806	의약품	21	책임	1176
801	기준	22	영향	1170
788	지속가능	23	의약품	1027
786	고객	24	ESG	1007
778	강화	25	프로세스	995
774	구성	26	거버넌스	974
751	품질	27	활동	949
739	리스크	28	사회	949
734	안전보건	29	보호	944
728	조직	30	배출량	942

다. ‘의약품’, ‘건강’ 등 제약 · 바이오업의 특징을 반영한 단어도 빈출 단어로 확인되었다.

한편 국내 제약 · 바이오사는 글로벌 선진 기업 대비 ‘안전’(6위)의 출현 빈도가 높게 확인되었다. 이는 우리나라의 사고 사망 만인율이 0.43으로 OECD 평균인 0.29를 크게 상회하는 만큼(김영태, 2022) 비교적 더 중요한 이슈로 다뤄지고 있다는 점과 함께 2021년 1월부터 시행된 중대재해처벌법이 영향을 미친 것으로 추정된다. ‘공급망’(11위), ‘리스크’(28위), ‘관리’(20위) 및 ‘교육’(5위), ‘지원’(16위) 등 공급망 내 리스크 요인에 대한 관리와 공급망 지원 관련 단어는 고르게 확인되었다. ‘제품’(12위), ‘사업’(19위), ‘고객’(24위), ‘품질’(27위) 등 비즈니스 관련 단어의 빈도도 높게 나타났다.

글로벌 선진 제약 · 바이오사는 업종의 본질적 목적 중 하나인 ‘건강’(1위)이 최고빈도 단어로 확인되었다. ‘사람’(6위), ‘환자’(13위), ‘접근성’(16위), ‘책임’(21위), ‘영향’(22위) 등의 단어 빈도도 높게 나타났다. 또한 ‘보고’(3위), ‘데이터’(19위), ‘ESG’(24위), ‘프로세스’(25위), ‘거버넌스’(26위), ‘배출량’(30위) 등도 고빈도 단어로 확인되었다.

단어 빈도 분석 결과 국내 기업에서는 안전, 공급망 관리 · 지원, 고객 만족, 비즈니스 관련 단어의 빈도가 30위권으로 확인되었다. 반면 글로벌 기업에서는 제약업의 비즈니스 목적과 의약 접근성 등 사회적 책임 관련 단어, ESG 거버넌스와

프로세스 수립, 기후 리스크 관리, 정보 공시 등과 관련한 단어가 높은 빈도로 확인되었다.

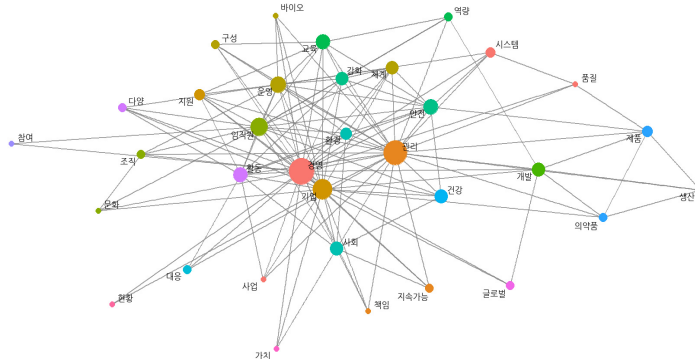
4.2 동시 출현 단어 분석

동시 출현 네트워크(Co-occurrence Network)는 동시에 출현한 단어의 관계를 네트워크 형태로 표현한 것으로, 단어들이 어떤 맥락에서 함께 사용되었는지 이해할 수 있게 해주는 시각화 방법이다. 네트워크 그래프는 단어를 나타내는 점 ‘노드(node)’와 연결된 노드를 잇는 선 ‘엣지(edge)’로 이루어져 있다.

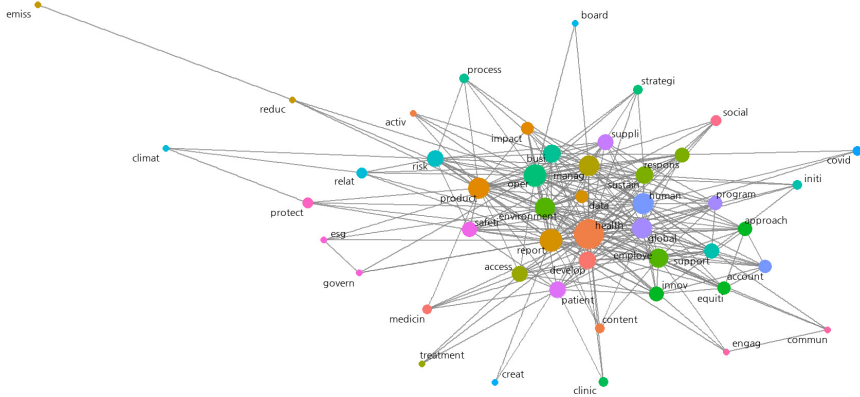
연결 중심성은 다른 노드들과 연결된 정도를 뜻하며, 커뮤니티는 단어 간 관계가 가까워 빈번하게 연결된 노드 집단을 의미한다(김영우, 2021). 연결 중심성은 키워드 노드 i 와 키워드 노드 j 가 이어져 있을 때 $a_{ij} = 1$, 그렇지 않을 때 $a_{ij} = 0$ 로 하여 <그림 4>와 같이 계산된다. n 은 전체 네트워크상의 키워드 노드 개수를 의미하며(이지원, 이향숙, 2022) 연결 중심성이 높은 단어는 문서의 핵심 주제로 간주된다(Ergun *et al.*, 2016).

$$C_D(i) = \sum_j^n a_{ij} / n - 1$$

<그림 4> 연결 중심성 산식



<그림 5> 국내 제약 · 바이오사 동시출현 네트워크 그래프



〈그림 6〉 글로벌 선진 제약·바이오사 동시출현 네트워크 그래프

본 연구에서는 한 문장 내에 함께 사용된 경우를 기준으로 동시 출현 네트워크를 도출하였다. 커뮤니티와 연결 중심성을 적용한 국내 기업의 동시 출현 네트워크를 <그림 5>, 글로벌 기업의 동시 출현 네트워크를 <그림 6>에 각각 나타내었다. 네트워크 그래프의 구조 파악을 용이하게 하기 위하여 연결 중심성에 따라 노드 크기를 다르게 하고, 커뮤니티별로 노드의 색상을 구분하였다.

국내 제약·바이오사의 텍스트 데이터는 ‘경영’, ‘관리’, ‘기업’ 등의 키워드를 중심으로 연결되어 있으며 ‘경영’·‘시스템’·‘품질’·‘사업’, ‘안전’·‘강화’, ‘임직원’·‘조직’·‘문화’, ‘건강’·‘의약품’·‘제품’ 등이 각각 동일 커뮤니티로 묶이는 것을 확인할 수 있었다(<그림 5>).

글로벌 제약·바이오사는 ‘건강’, ‘보고’, ‘기업’, ‘사람’ 등의 키워드를 중심으로 연결되어 있으며 ‘데이터’·‘보고’, ‘지속가능’·‘책임’, ‘약’·‘개발’, ‘배출’·‘감축’ 등이 각각 동일 커뮤니티로 묶이는 것을 확인하였다(<그림 6>).

4.2.1 EU 기업 지속가능성 실사 지침 주요 단어 기준 분석

제약·바이오 업종의 두 연구 대상 그룹이 CS3D의 주요 키워드에 대해 어떤 활동을 추진하고 있는지 파악하고자 CS3D 핵심 키워드 기준 동

시 출현 단어를 확인하였다. CS3D 핵심 키워드는 한가록 등(2022)의 선행 연구를 바탕으로 선정하였다. 해당 연구는 파이썬을 이용해 2022년 2월 23일 공개된 CS3D 원문의 키워드를 분석하여 지침 원문의 키워드 연결 중심성 결과를 <표 9>와 같이 도출하였다(한가록 등, 2022).

〈표 9〉 EU 기업 지속가능성 실사 지침 원문 연결 중심성 분석 결과

연번	단어	중심성
1	부정적 영향	27.014
2	기업실사	20.066
3	지속가능성	16.987
4	글로벌	11.242
5	환경	8.090
6	법	7.028
7	자회사	6.666
8	EU	6.013
9	인권	5.487
10	기후	4.137

CS3D 원문 단어 중 연결 중심성이 높게 나온 ‘부정적 영향’, ‘실사’, ‘지속가능성’, ‘환경’, ‘인권’, ‘기후’, 그리고 CS3D의 제정 목적을 고려해 ‘공급망’을 분석 기준 단어로 설정하여 동시 출현 단어를 비교 분석하였다.

CS3D는 공급망에서의 환경·인권 관련 실질적·잠재적 부정 영향을 식별해 보고하도록 규정한다. ‘부정적 영향’을 기준으로 두 연구 대상 그룹 지속가능경영보고서 내 동시출현 단어를 분석한 결과, 국내 기업은 해당 단어가 확인되지 않았다. 이에 ‘영향’을 기준 단어로 하여 분석을 수행하였다(<표 10>).

국내 기업 지속가능경영보고서에서 ‘영향(impact)’과 함께 출현한 주요 단어를 살펴보면 ‘환경’(1위), ‘사회’(10위) 등 ESG 영향과 ‘경영’(2위), ‘제품’(5위), ‘기업’(8위) 등 기업 운영 전반에 대한 영향을 중요하게 다루고 있음을 알 수 있다.

글로벌 제약·바이오사 지속가능경영보고서에서 ‘부정적 영향(adverse impact/negative impact)’과 함께 사용된 단어를 살펴보면 부정적 영향에 대해 어떻게 ‘접근’(2위)하고 ‘평가’(8위)하며 ‘보고’(9위)할 것인지와 관련한 ‘프로세스’(10위)를 갖추고 있음을 유추할 수 있다. 이는 CS3D에서 요구하는 부정적 영향 식별 및 관리 프로세스를 어느 정도 내재화하고 있음을 나타낸다.

<표 10> CS3D 주요 단어 기준 제약·바이오 업종 동시 출현 단어 분석(부정적 영향)

(부정적)영향		
국내 기업	순위	글로벌 기업
환경	1	기업
경영	2	접근
활동	3	사람
관리	4	건강
제품	5	실적
최소화	6	관리
변화	7	경제
기업	8	평가
임직원	9	보고
사회	10	프로세스

CS3D에서는 공급망 내 환경·인권 영향을 방지하기 위한 실사 정책 수립 및 내재화, 지속 가능한 공급망 관리 프로세스 구축을 요구하고 있다. 이에 ‘실사(due diligence)’, ‘지속가능(sustainability)’과

동시 출현한 단어를 분석하였다(<표 11>).

국내 제약·바이오사에서는 CS3D의 제정 목적인 ‘공급망’ 실사 관련 단어가 10위권 내 주요 단어로 확인되지 않았다. 반면 글로벌 제약·바이오사에서는 ‘공급망’(5위)의 ‘인권’(1위) ‘리스크’(4위), 관련 ‘프로세스’(3위) 수립과 ‘보고’(7위) 등의 단어가 함께 출현하고 있음을 확인할 수 있었다. 이를 통해 CS3D가 요구하는 공급망 리스크 실사 프로세스 수립 역량에서 두 그룹 간 차이를 유추할 수 있다.

‘지속가능(sustainability)’을 기준어로 분석한 결과, 글로벌 기업에서는 ‘공급망’(9위)이 고빈도로 동시 출현함으로써 지속 가능한 공급망 관리를 주요 이슈로 설정하고 있는 것을 알 수 있다(<표 11>).

<표 11> CS3D 주요 단어 기준 제약·바이오 업종 동시 출현 단어 분석(실사/지속가능)

실사		
국내 기업	순위	글로벌 기업
관리	1	인권
품질	2	사업
제품	3	프로세스
경영	4	리스크
리스크	5	공급망
안전	6	접근성
진행	7	보고
기업	8	접근
인권	9	지속가능
개선	10	프로그램
지속가능		
국내 기업	순위	글로벌 기업
경영	1	보고
보고	2	환경
기업	3	기업
사회	4	건강
관리	5	전략
글로벌	6	관리
활동	7	개발
산업	8	수행
이해관계	9	공급망
체계	10	사업

CS3D에서는 ‘환경’ 및 ‘인권’ 관련 부정적 영향 파악 및 완화를 요구하고 있다. 이에 해당 단어들을 기준으로 두 그룹 지속가능경영보고서 내 동시 출현 단어를 분석하였다(<표 12>).

‘환경(environment)’을 기준으로 한 분석 결과에서 두 그룹 모두 ‘영향’이라는 단어가 확인되어 CS3D에서 요구하는 환경 관련 영향을 파악하고 있음을 유추할 수 있다. 글로벌 제약·바이오사의 경우 ‘보고’(6위) 단어가 확인되어 환경 관련 정보 공시에 보다 적극적인 것을 유추할 수 있다(<표 12>).

<표 12> CS3D 주요 단어 기준 제약·바이오 업종 동시 출현 단어 분석(환경/인권)

환경		
국내 기업	순위	글로벌 기업
안전	1	건강
경영	2	안전
관리	3	사람
사회	4	지속가능
체계	5	보호
운영	6	보고
활동	7	기업
임직원	8	글로벌
영향	9	영향
기업	10	제품
인권		
국내 기업	순위	글로벌 기업
경영	1	리스크
임직원	2	관리
사회	3	공급망
관리	4	임직원
교육	5	존중
기업	6	지속가능
활동	7	글로벌
안전	8	보고
공헌	9	표준
지속가능	10	정책

‘인권(human rights)’을 기준으로 분석한 결과, 글로벌 기업의 분석 결과에서 CS3D가 규정하는

‘글로벌’(7위) ‘공급망’(3위) 내 ‘리스크’(1위) ‘관리’(2위)가 높은 빈도의 동시 출현 단어로 등장한 것을 알 수 있다. 특히 ‘글로벌’(7위) ‘표준’(9위) 참조와 인권 관련 ‘정책’(10위) 마련, 인권 실사 내용에 대한 ‘보고’(8위) 등 CS3D가 공급망 인권 실사에서 요구하는 전반적인 사항을 잘 반영하고 있는 것으로 추정된다(<표 12>).

CS3D 제15조에서는 기후변화 대응 요구를 명시하고 있다. 또한 지침 전반에 걸쳐 공급망 리스크 실사 체계 구축과 이행을 규정하고 있다. 이에 ‘기후’와 ‘공급망’ 두 단어를 기준으로 두 그룹의 동시 출현 단어를 분석하였다(<표 13>).

<표 13> CS3D 주요 단어 기준 제약·바이오 업종 동시 출현 단어 분석(기후/공급망)

기후		
국내 기업	순위	글로벌 기업
변화	1	변화
대응	2	보호
관리	3	리스크
환경	4	기업
경영	5	관리
위험	6	영향
온실가스	7	보고
전략	8	배출
활동	9	감축
기회	10	제품
공급망		
국내 기업	순위	글로벌 기업
관리	1	지속가능
경영	2	제품
임직원	3	관리
안전	4	리스크
활동	5	글로벌
지원	6	다양성
운영	7	인권
리스크	8	프로그램
평가	9	보고
사회	10	환경

‘기후(climate)’를 기준으로 분석한 결과, 국내 제약·바이오사는 기후 이슈와 관련해 ‘온실가스’(7위)를 비중 있게 언급하고 있으나, 글로벌 제약·바이오사에서 확인된 ‘배출량’(8위), ‘감축’(9위) 및 ‘보고’(7위) 등의 단어는 나타나지 않았다(<표 13>).

‘공급망(supply chain)’과 관련해 국내 기업에서는 ‘지원’(6위)이 동시 출현하고 있으나, ‘안전’(4위) 외에 구체적으로 어떤 리스크에 초점을 맞춰 공급망을 관리하고 있는지 명확하게 확인되지 않는다. 글로벌 제약·바이오사에서는 ‘글로벌’(5위) 공급망의 ‘인권’(7위) 및 ‘다양성’(6위), ‘환경’(10위) 등 구체적인 ‘리스크’(4위)를 ‘관리’(3위)하고 ‘보고’(9위)하고 있는 것으로 추정할 수 있다(<표 13>).

4.2.2 기후변화 관련 단어 기준 분석

CS3D에서는 환경에 대한 부정적 영향을 ‘국제 환경 협약에 따른 금지 의무를 위반해 발생하는 환경에 대한 악영향’으로 규정하고 있다(European Commission, 2022). 영향 판단 기준이 되는 국제협약에는 유럽기후법, Fit for 55 등 EU의 기후법안이 포함된다.

이에 본 연구에서는 ‘온실가스(greenhouse gas)’, ‘탄소(carbon)’를 기준 단어로 설정하고 두 연구 대상 그룹 간 지속가능경영보고서 텍스트 내에서 해당 기준 단어와 동시 출현한 단어의 빈도를 <표 14>와 같이 분석하였다.

‘온실가스’는 두 그룹 모두 ‘배출량’, ‘감축’과 ‘에너지’ 등이 주요 동시 출현 단어로 확인된다. 국내 기업 대비 글로벌 제약·바이오사의 특징은 ‘데이터’(7위), ‘보고’(4위) 단어가 확인된다는 것이다. 앞선 분석 결과와 유사하게 글로벌 기업이 데이터 관리와 공시를 보다 적극적으로 이행하고 있음을 보여준다.

‘탄소’는 두 그룹 모두 ‘탄소중립’과 ‘배출량’ 등이 주요 단어로 확인된다. 다만 글로벌 제약·바이오사는 ‘달성’(8위)이 추가로 확인되어 탄소

중립 달성에 비교적 강한 의지를 보이고 있음을 알 수 있다. 특히 글로벌 기업의 경우 ‘탄소발자국’(9위)과 ‘공급망’(10위)이 10위권에 포함되어 제품 탄소 배출량에 비교적 더 높은 관심을 갖고 있는 것으로 보인다. 이는 글로벌 기업이 CS3D에서 요구하는 공급망의 기후 및 환경 영향 관리 측면에서 국내 기업에 비해 보다 진일보한 대응체계를 갖추고 있음을 보여준다.

<표 14> 기후 관련 제약·바이오 업종 동시 출현 단어 분석(온실가스/탄소)

온실가스		
국내 기업	순위	글로벌 기업
배출량	1	배출
에너지	2	스코프
관리	3	감축
검증	4	보고
환경	5	에너지
사용량	6	기후
활동	7	데이터
변화	8	환경
감축	9	프로토콜
대응	10	탄소

탄소		
국내 기업	순위	글로벌 기업
배출량	1	배출
중립	2	감축
기후	3	중립
저감	4	환경
변화	5	기후
활동	6	지속가능
관리	7	스코프
에너지	8	달성
사용	9	발자국
대응	10	공급망

4.2.3 인권 관련 단어 기준 분석

CS3D에서는 인권에 대한 부정적 영향을 ‘국제협약이 제시하는 권리 또는 금지 사항을 위반해 보호 대상자에게 미치는 악영향’으로 정의하고

있다. 이때 기준이 되는 국제협약은 「UN 국제인권협약」, 「세계인권선언」, 「아동권리협약」, 「국제노동법」 등이다(European Commission, 2022). 이에 ‘차별’, ‘강제’, ‘아동’, ‘노예’를 인권 관련 주요어로 설정해 동시 출현 단어를 분석하였다(<표 15>, <표 16>).

<표 15> 인권 관련 제약·바이오 업종 동시 출현 단어 분석(차별/강제)

차별		
국내 기업	순위	글로벌 기업
임직원	1	임직원
금지	2	사람
교육	3	괴롭힘, 희롱
사회	4	다양성
다양	5	표준
인권	6	사회
성별	7	보고
공급망	8	노동
조치	9	이니셔티브
업무	10	성별

강제		
국내 기업	순위	글로벌 기업
금지	1	보고
근로	2	노동
노동	3	리스크
경영	4	사람
공개	5	글로벌
기업	6	사업
안전	7	공급망
아동	8	아동
바이오	9	GRI
근로자	10	정책

‘차별(discrimination)’에 대한 분석 결과를 보면 두 그룹 모두 ‘임직원’과 ‘성별’에 대한 차별과 ‘다양성’ 이슈를 언급하고 있음을 알 수 있다(<표 15>). 글로벌 기업에서는 ‘괴롭힘’(3위) 등 구체적인 차별 행위를 언급하고 있으며, 차별 관

련 ‘표준’(5위)과 ‘이니셔티브’(9위)를 지지하고 관련 내용을 ‘보고’(7위)하는 등 보다 체계적이고 적극적으로 차별 이슈를 다루고 있음을 확인할 수 있다.

‘강제(forced)’와 관련해서는 두 그룹 모두 ‘노동’과 ‘아동’ 단어가 동시 출현하고 있으며, 국내 기업에서는 ‘공개’, 글로벌 기업에서는 ‘보고’ 단어가 동시에 나타남에 따라 두 그룹 모두 정보 공개와 보고의 필요성을 인식하고 있는 것으로 보인다(<표 15>). 글로벌 기업에서는 ‘공급망’(7위)에서의 강제노동 ‘리스크’(3위)를 관리하며 관련 ‘정책’(10위)을 수립하고 있는 등 CS3D가 요구하는 인권 리스크 관리 정책을 비교적 내재화하고 있는 것으로 파악된다.

‘아동(child)’ 이슈의 경우 국내 기업은 ‘취약계층’(10위), ‘청소년’(8위), ‘지원’(1위), ‘사회’(3위), ‘공헌’(6위) 등 사회공헌 차원에서 다루고 있는 것으로 보인다(<표 16>). 글로벌 기업에서 ‘아동’과 가장 자주 사용된 단어는 ‘노동’(1위)으로 아동에 대한 노동 이슈를 중요하게 다루고 있는데, 특히 CS3D에서 요구하는 ‘공급망’(6위), ‘인권’(8위) 리스크로서 아동 노동 이슈를 관리하고 있음을 확인할 수 있다.

‘노예(slavery)’의 경우 국내 제약·바이오사에서는 유의미한($n \geq 2$) 동시 출현 단어가 확인되지 않았다(<표 16>). 글로벌 기업에서는 ‘노예’가 ‘공급망’(6위) 및 ‘인권’(2위)과 함께 나타났는데 ‘인권’(2위) 이슈를 ‘관리’(7위)하고, 관련 ‘원칙’(9위)과 ‘정책’(10위)을 제정하며, 활동을 ‘보고’(8위)하는 등 CS3D가 요구하는 공급망 내 강제노동 근절을 위한 체계를 갖추고 있음을 유추할 수 있다.

동시 출현 단어 분석 결과를 종합적으로 살펴보면 국내 제약·바이오 기업은 글로벌 기업 대비 부정적 이슈의 식별과 보고, 공급망 실사 이행과 리스크 관리 프로세스가 미흡하며 환경, 기후, 인권 이슈에 대한 정책 수립과 공시 노력이 비교적 부족한 것으로 확인되었다.

〈표 16〉 인권 관련 제약·바이오 업종 동시 출현 단어 분석(아동/노예)

아동		
국내 기업	순위	글로벌 기업
지원	1	노동
활동	2	건강
사회	3	지원
임직원	4	프로그램
지역	5	돌봄
공헌	6	공급망
지역사회	7	치료
청소년	8	인권
기업	9	개선
취약	10	이니셔티브

노예		
국내 기업	순위	글로벌 기업
없음	1	현대
	2	인권
	3	성명서
	4	지원
	5	교육
	6	공급망
	7	관리
	8	보고
	9	원칙
	10	정책

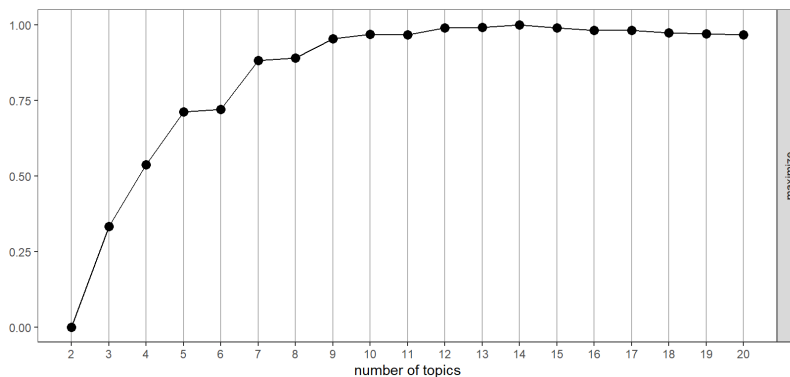
4.3 토픽 모델링 분석

토픽 모델링 분석 방법은 여러 가지가 존재하나 본 연구에서는 가장 대표적인 방법인 Latent Dirichlet Allocation(이하 LDA) 방법을 사용하였다(Blei *et al.*, 2003). LDA는 사용자의 경험을 분석하고 관리하거나(연다인 등, 2020), 머신러닝과 결합해 사기 위험을 탐지하는(이동우, 민진영, 2021) 등 다양한 확장성을 가진 토픽 모델링 방법이다.

본 연구는 Griffiths2004를 활용한 하이퍼파라미터 튜닝으로 연구 대상 그룹별 문서의 최적 토픽 수를 도출하였다. Griffiths2004는 Griffiths와 Steyvers(2004)가 제안한 복잡도(perplexity) 기반 모델 설명력 평가 지표이다. 다양한 연구에서 토픽모델의 성능을 측정하는 데 활용되었으며 값이 클수록 모델의 설명력이 높다(김영우, 2021).

본 연구에서는 토픽별 상위 10개 단어를 중심으로 각 토픽의 의미를 추론하였다. 고빈도 단어는 여러 토픽에 중복 분류될 수 있으므로 명확한 주제 분류를 위해 빈도 분석에서 도출된 상위 5개 단어(국내: ‘기업’, ‘경영’, ‘관리’, ‘임직원’, ‘활동’/글로벌: ‘기업’, ‘환경’, ‘관리’, ‘제품’, ‘보고’)와 지속가능경영보고서 관련 단어(‘지속가능’, ‘보고서’)를 삭제한 후에 분석하였다.

국내 제약·바이오사의 최적 토픽 수를 도출하기 위해 <그림 7>과 같이 Griffiths2004 성능지표를



〈그림 7〉 Griffiths2004 성능지표에 따른 국내 제약·바이오사 최적 토픽 수(k=7)

활용하였다. 그림에서 x축은 토픽 수, y축은 성능 지표를 의미하며 이때 성능 지표는 0~1로 정규화한 값이다. y축의 값이 클수록 모델의 설명력이 높음을 의미한다. <그림 7>을 보면 토픽 수가 7이 될 때까지 성능 지표가 가파르게 증가하고, 포화된 이후로는 일정한 값에 수렴하는 모습이다. 성능이 비슷하다면 보다 단순한 모델을 사용하는 것이 효과적이므로 토픽 수를 7개로 선정하였다. 이처럼 토픽 수를 늘려도 성능이 크게 향상되지 않고 등락을 반복하기 시작하는 지점이 토픽 수를 결정하는 구간이 된다(김영우, 2021).

토픽 수 k=7로 설정해 LDA 분석을 수행한 결과, 국내 제약·바이오사 지속가능경영보고서 텍스트의 7개 토픽별 주요 단어는 <표 17>과 같이 확인되었다.

7개 토픽별 주요 단어를 기반으로 해석한 각

토픽의 의미는 <표 18>과 같이 나타낼 수 있다.

국내 기업 LDA 분석 결과 ‘사업장 에너지 사용량 및 폐기물 배출’ 관련 주제가 도출된 것은 국내 제약·바이오사가 실사 규제에 적절하게 대응하고 있음을 보여준다(<표 18>). 에너지 사용으로 인한 온실가스 배출, 폐기물 등의 오염물질 배출 항목은 공급망 실사의 주요 평가항목으로서 적극 대응해야 하는 이슈이기 때문이다.

‘헬스케어 산업 및 환경’과 ‘건강기능 식품’ 등 사업전략 관련 주제, ‘신약 연구개발’ 주제는 국내 제약·바이오 기업의 비즈니스 역량 강화 노력을 보여준다.

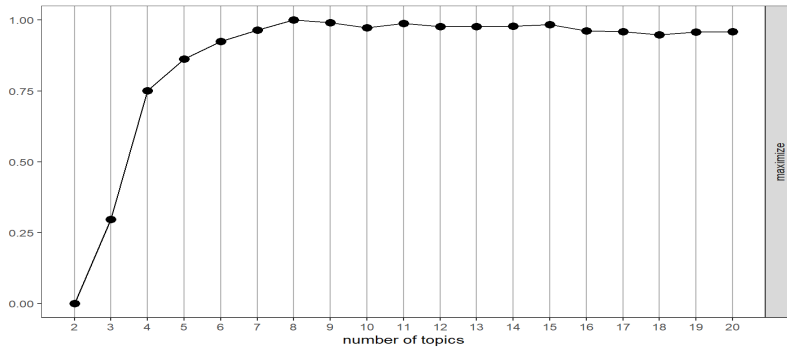
CS3D가 요구하는 ‘공급망’ 이슈의 경우 중요 주제로 자리 잡고 있음을 확인하였으나 공급망에 대한 평가나 리스크 실사보다는 ‘지원’에 보다 초점을 맞추고 있음을 알 수 있다.

<표 17> 국내 제약·바이오사 LDA 분석 토픽별 주요 단어

순위	토픽 1	토픽 2	토픽 3	토픽 4	토픽 5	토픽 6	토픽 7
1	제약	케어	건강	진단	안전	바이오	신약
2	사업장	헬스	식품	책임	운영	백신	건강
3	에너지	산업	기능	사회	활동	구성	의약품
4	폐기물	보고	현황	분자	개발	품질	품질
5	사용량	환경	인증	추진	사회	모니터링	개선
6	현황	협회	검증	가치	제품	공급망	성실
7	신약	비율	글로벌	검사	공급망	대응	글로벌
8	안전보건	생태계	전략	합계	지원	임상시험	국민
9	배출	연결	주요	매출액	체계	이행	연구
10	목표	품질	헬스	영업이익	환경	영향	우수

<표 18> 국내 제약·바이오사 토픽명 및 주요 단어

토픽	주요 단어	토픽명
1	사업장, 에너지, 폐기물, 사용량, 배출, 목표	사업장 에너지 사용량 및 폐기물 배출
2	케어, 헬스, 산업, 환경, 생태계, 품질	헬스케어 산업 및 환경
3	건강, 식품, 기능, 글로벌, 전략, 주요	건강기능식품 및 주요 사업전략
4	진단, 책임, 사회, 가치, 검사, 매출액, 영업이익	진단검사 투자 및 사회책임
5	안전, 운영, 활동, 공급망, 지원, 체계, 환경	공급망 안전·환경관리 및 지원
6	백신, 품질, 모니터링, 대응, 임상시험, 영향	백신 개발
7	신약, 의약품, 품질, 개선, 글로벌, 연구, 우수	신약 연구개발



<그림 8> Griffiths2004 성능지표에 따른 글로벌 제약·바이오사 최적 토픽 수(k=8)

글로벌 제약·바이오사의 LDA 분석을 위해 Griffiths2004 모델 성능 지표를 확인한 결과, 토픽 수 k는 8개가 될 때까지 지속 증가하다가 해당 지점에서 정점을 찍은 후 다시 하락한다(<그림 8>). 이에 토픽 수 k=8로 설정해 분석을 수행하였다. 글로벌 제약·바이오사 지속가능경영보고서의 8개 토픽별 주요 단어는 <표 19>와 같이 확인되었으며, 각 토픽별 의미는 <표 20>과 같이 정리할 수 있다.

글로벌 제약·바이오사의 경우 ‘글로벌 공급망 책임관리’ 주제 내에 ‘리스크’, ‘책임’ 등의 단어를 포함하고 있어 공급망에 대한 지원을 넘어 리스크 관리까지 이행하고 있음을 알 수 있다(<표 20>).

‘기후환경 보호 및 농작물 조달’ 주제도 확인되

었는데, 농작물 등 식물원료는 의약품의 주요 원재료로서 제약·바이오 산업의 원료 수급 안정성과 직결하는 이슈다. 이는 글로벌 제약·바이오사가 기후위기로 인한 가치사슬 전반의 비즈니스 영향을 고려하고 있음을 보여주며, 국내 기업의 기후변화 이슈가 사업장 단위의 에너지 사용 및 폐기물 배출에 집중되어 있던 것과 차이를 보인다.

또한 ‘다양성’, ‘불공평’ 등의 단어가 포함된 ‘ESG 강화 및 GRI 공시’ 주제를 통해 글로벌 제약·바이오사가 인권의 주요 개념인 DEI(Diversity, Equity and Inclusion), 즉 조직의 다양성·형평성·포용성을 적극 고려하고 있으며 관련 활동을 GRI 인덱스 등 글로벌 공시 기준에 따라 보고하고 있음을 알 수 있다.

<표 19> 글로벌 제약·바이오사 LDA 분석 토픽별 주요 단어

순위	토픽 1	토픽 2	토픽 3	토픽 4	토픽 5	토픽 6	토픽 7	토픽 8
1	보호	리스크	재정	환자	ESG	백신	공급망	테스트
2	작물	현장	창출	사회	강화	ESG	환경	암
3	안전	안전	사회	접근권	공동체	혁신	개발	환자
4	기후	프로세스	자금	리스크	창출	글로벌	글로벌	코로나19
5	권리	컴플라이언스	관리자	코로나19	다양성	불공평	프로그램	진단
6	환경	내부	암	연계	불공평	책임	리스크	아시아
7	조치	요구	개선	약	GRI	코로나19	데이터	태평양
8	조달	권리	담당자	치료	환자	진진	지원	프로그램
9	GRI	문서	연구	질병	프로그램	목표	접근성	질병
10	농부	표준	재무	영향	인덱스	소비	책임	영향

<표 20> 글로벌 제약·바이오사 토픽명 및 주요 단어

토픽	주요 단어	토픽명
1	보호, 작물, 기후, 환경, 조치, 조달, 농부	기후환경 보호 및 농작물 조달
2	리스크, 현장, 안전, 프로세스, 컴플라이언스, 문서, 표준	현장 및 공정 안전을 위한 국제 기준 준수
3	재정, 창출, 사회, 자금, 암, 개선, 연구, 재무	암 대응 연구개발 투자 및 사회 가치 창출
4	환자, 사회, 접근권, 코로나19, 약, 치료, 질병, 영향	질병 치료 및 사회 영향
5	강화, 공동체, 다양성, 불공평, GRI, 프로그램, 인덱스	ESG 강화 및 GRI 공시
6	백신, 혁신, 글로벌, 불공평, 책임, 코로나19, 진전, 소비	질병 백신 연구개발 및 약물 접근성에 대한 책임
7	공급망, 환경, 개발, 글로벌, 프로그램, 리스크, 데이터, 지원, 접근성, 책임	글로벌 공급망 책임 관리
8	테스트, 암, 환자, 코로나19, 진단, 질병	질병 진단 제품 및 서비스

<표 21> 주요 분석결과

단어빈도 분석	<ul style="list-style-type: none"> - 국내 제약·바이오사는 안전, 공급망 리스크 관리 및 지원, 비즈니스 관련 단어 빈도 높음 - 글로벌 제약·바이오사는 기업의 목적, 사회적 책임과 데이터 공시, 거버넌스, 배출량 등의 단어 빈도가 국내 제약·바이오사 대비 높음
동시출현 단어분석	<p>CS3D 주요 단어 기준</p> <ul style="list-style-type: none"> - 부정적 영향 관련 접근·평가·보고·프로세스 단어가 확인되는 글로벌 기업 대비 국내 제약·바이오사의 경우 부정적 영향 관련 단어 보고 부재 - 국내 대비 글로벌 제약·바이오사의 경우 구체적인 공급망 리스크의 종류 및 관리 의지 실행을 위한 프로세스 수립, 정보공시 관련 단어 빈출 <p>기후 관련 주요 단어 기준</p> <ul style="list-style-type: none"> - 글로벌 제약·바이오사의 경우 제품 탄소발자국 및 공급망 탄소 배출량 데이터 관리, 보고 등의 단어 빈출 <p>인권 관련 주요 단어 기준</p> <ul style="list-style-type: none"> - 글로벌 제약·바이오사는 공급망 강제노동 관리, 관련 원칙, 정책 수립, 정보공시 관련 단어 빈출 - 국내 제약·바이오사는 공급망 내 아동 노동, 노예 노동 관련 내용 공시 미흡
토픽 모델링 분석(LDA)	<ul style="list-style-type: none"> - 국내 제약·바이오사의 경우 ‘에너지 사용량 및 폐기물 배출’, ‘사업 전략’, ‘신약 연구개발’, ‘공급망 관리’ 등의 토픽이 확인되나 공급망 ‘실사’보다 ‘지원’에 초점 - 글로벌 제약·바이오사는 ‘공급망 리스크 관리 및 책임’, ‘기후환경 보호 및 농작물 조달’, ‘ESG 강화 및 GRI 공시’ 등의 토픽 도출

단어 빈도 분석, 동시 출현 단어 분석, LDA에 의한 주요 분석결과를 위의 <표 21>과 같이 정리하였다.

V. 결 론

EU는 다양한 환경 규제와 ESG 공시 제도를 강화해 비재무적 무력 장벽을 높이며 국내 수출기업에 부담을 가중하고 있다. ESG 규제화는 개별 기

업을 넘어 기업의 자회사와 공급망 전반까지 확대 적용되는 추세이다. 각국에서 공급망 인권·환경 실사 법제화 움직임을 확대하고 있으며, 그중에서도 특히 EU의 CS3D는 향후 개별 회원국이 자국법으로 제정 시 우리나라의 EU 수출 기업에 큰 영향을 미칠 수밖에 없다.

이에 본 연구에서는 국내 제약·바이오 기업이 EU의 CS3D 직·간접 적용에 대비해 공급망 리스크와 자사의 인권·환경 경영을 잘 이행하고 있는

지 현황을 살펴보았다. 또한, 인권·환경 경영을 잘 수행하고 있다고 평가받는 글로벌 선진기업의 경영 현황과 비교 분석하여 CS3D에 대응하기 위한 효과적인 벤치마킹 방안을 도출하고자 하였다.

본 연구 결과는 다음과 같은 시사점을 제공한다. 첫째, 국내 제약·바이오 기업은 CS3D에서 요구하는 환경·인권 관련 부정적 이슈를 식별 및 보고하기 위해 노력을 기울여야 한다. ‘부정적 영향’을 기준으로 동시 출현 단어를 분석한 결과 국내 기업은 어떠한 단어도 확인되지 않은 반면, 글로벌 선진 기업은 ‘접근’, ‘관리’, ‘평가’, ‘보고’ 등 부정적 영향을 관리하고 보고하는 노력을 유추할 수 있는 단어들이 확인되었다. 특히 국내 기업의 경우 노예 노동, 아동 노동 등과 관련해 유의미한 동시 출현 단어가 거의 확인되지 않았는데, 해당 주제의 경우 국내에서는 위험성이 비교적 적다 할지라도 글로벌 공급망으로 확장하면 매우 민감하고 중요한 이슈가 될 수 있으므로 이에 대한 관리와 정보 공시를 이행하여야 한다.

둘째, 국내 제약·바이오 기업은 CS3D에서 요구하는 구체적인 공급망 실사 이행 지침을 수립하고 내재화하여야 한다. CS3D를 비롯한 주요 공급망 실사 프레임워크에서는 ‘위험의 가시성’을 핵심적으로 요구한다. 복잡하게 얽힌 공급 사슬에서 위험을 식별하기 위해서는 이를 확인하기 위한 체계적인 프로세스가 필요하다. 텍스트 마이닝 분석 결과, 글로벌 기업에서는 ESG 경영 이행을 위한 기업의 ‘정책’과 ‘프로세스’ 관련 단어가 지속적으로 확인되었다. ‘표준’과 ‘이니셔티브’가 고빈도 단어로 등장한 것에도 주목할 필요가 있다. 환경·인권 관련 국제 표준이나 글로벌 이니셔티브는 관련 경영 체계 구축에 큰 도움이 될 수 있으므로 국내 제약·바이오 기업은 이를 적극 활용해 기업 내 공급망 위험 관리 체계를 내재화하려는 노력을 가일층 기울여야 한다.

셋째, 국내 제약·바이오 기업은 CS3D 대응을 위해 자사 공급망을 지원의 대상이 아닌 리스크 평가 대상으로서 관리할 필요가 있다. ‘공급망’ 단

어와 동시 출현한 단어 및 LDA 토픽 모델링 분석 결과 국내 기업은 글로벌 기업 대비 ‘지원’, ‘성장’, ‘교육’ 등의 단어를 사용하는 반면, 글로벌 기업은 ‘인권’, ‘리스크’, ‘정책’ 등의 단어를 함께 사용하고 있음을 확인하였다. 이는 공급망에 대한 두 연구대상의 접근 방식 차이를 보여준다. 자원이 부족한 중소기업의 경우 ESG 역량 강화를 위해 대기업의 지원이 필요하다. 그러나 공급망 관리가 규제화하고 있는 현재와 같은 상황에서는 단순 지원을 넘어 중소기업 공급망이 가진 리스크를 우선적으로 파악하고, 이를 저감하기 위한 방향성을 견지한 채 차별화된 지원을 제공하여야 한다.

넷째, 국내 제약·바이오 기업은 CS3D에서 요구하는 환경·인권 정보 공시를 위해 ESG 데이터 관리를 고도화하여야 한다. 국내 기업 대비 글로벌 기업 분석 결과에서 일관되게 ‘보고’와 ‘데이터’ 단어가 확인되었다. 데이터 관리는 공급망 실사 내용을 직접 공시하여야 하는 CS3D 직접 적용 기업에도, 원청사의 실사 요구에 대응해야 하는 간접 적용 기업인 협력사에도 중요한 이슈이므로 관련 역량 강화가 필요하다.

다섯째, 정부는 우리 기업의 공급망 실사 대응 역량 강화를 위해 컨설팅, 재정 지원 등과 함께 공급망 환경·인권 관리 의무화 제도 수립을 추진하여야 한다. 현재 글로벌 국가들의 공급망 관련 법 제정 현황이나 EU 의회의 CS3D 지침 초안 수정 추세를 보면 그 기준과 범위가 갈수록 강화 및 확대되어 가고 있음을 알 수 있다. ESG 환경에서의 공급망 실사법을 연구한 장혜진, 최윤정(2023)이 제안하였듯이 현재의 글로벌 흐름에 발맞추고 우리 기업이 이에 적기 대응하기 위해서는 적절한 국내법 제정이 필요할 것으로 보인다. 향후 금융위원회, 금융감독원, 한국거래소에서 제정할 비재무정보 공시 규정에 공급망 실사 내용 공개를 명시하거나, 현재 국회에 계류 중인 「경제 안보를 위한 공급망 안정화 지원 기본법안」 등에 이러한 내용을 포함하는 것이 바람직하다.

본 연구의 학술적 의의는 다음과 같이 정리할

수 있다. 첫째, 본 연구는 대응 중요성과 시급성이 높은 EU CS3D에 대한 국내 제약·바이오사 기업의 대응 현황을 분석하고 개선 방향을 제안한 최초의 연구라는 점에서 의의를 갖는다. 향후 다른 업종별 연구의 기초 자료로 활용될 수 있을 것으로 기대한다.

둘째, 정성적·비정형적 데이터인 기업의 지속가능경영보고서를 텍스트 마이닝으로 정량 분석함으로써 CS3D와 같은 신규 규제에 대한 기업의 대응 현황 파악 및 구체적 대응 방안 도출을 실증하였다. 특히 글로벌 선진 제약·바이오 기업 현황과의 비교 분석을 통하여 기업 경영의 벤치마킹 수단으로서 텍스트 마이닝의 효용성을 증명하였다.

셋째, 동시 출현 단어 분석 및 LDA 토픽 모델링을 적용해 특정 규제에 대한 기업의 대응 방향은 물론 그 수준을 분석할 수 있다는 점을 확인하였다. 앞으로도 ESG 관련 규제는 지속 증가하고, 다양한 요인을 동시에 실행하도록 요구하는 방식으로 복잡화할 전망이다. 이를 효과적으로 분석하는 방법론으로서 동시 출현 단어 분석과 LDA 토픽 모델링이 확대 적용될 것으로 기대한다.

다만 본 연구는 제약·바이오 기업의 환경·인권 경영 내용을 개괄적, 포괄적으로 다루었기 때문에 기업별 경영 체계와 현황을 세부적으로 분석하지는 못하였다. 텍스트 마이닝이라는 연구 방법의 특성상 연구 대상의 ESG 경영 및 EU CS3D 대응 수준을 객관적으로 등급화하기보다는 각 기업이 주요 주제를 다루는 관점과 적극성, 비중의 차이를 확인하는 데 그친 점도 한계로 지적된다. 또한, 기업의 환경·인권 경영 현황을 망라한 지속가능경영보고서의 경우 대표성과 비교 가능성에서 가장 적절한 분석데이터였음에도 불구하고 별도의 환경 및 인권 보고서를 발간하는 기업이 있는 점을 고려할 때 이를 더 포함해 분석하지 못하였다는 한계가 있다. 향후 연구에서는 분쟁광물 보고서, 기후행동 보고서 등 데이터 범위를 확대해 더욱 폭넓은 분석이 이루어져야 할 것이다. 추가로, 연구 대상인 국내 기업의 EU 법인 설립 여부

및 EU 내 매출액 확인이 어려워 CS3D 직접 적용 대상인 대기업과 간접 적용 대상인 중견·중소기업을 나누어 분석하지 못한 것도 향후 연구에서 보완하여야 할 점이다.

참고 문헌

- [1] 강노경, *EU 공급망실사지침(안) 주요내용과 기업사례*, 한국무역협회, 2022.
- [2] 관계부처 합동, *K-ESG 가이드라인 v 1.0*, 2021.
- [3] 김성근, 조혁준, 강주영, “학술연구에서의 텍스트 마이닝 활용 현황과 주요분석기법”, *정보기술아키텍처연구*, 제13권, 제2호, 2016, pp. 317-329.
- [4] 김영우, *쉽게 배우는 R 텍스트 마이닝*, 이지스퍼블리싱, 2021.
- [5] 김영태, “당정, 2026년까지 OECD 수준으로 중대재해 감축”, *경북매일*, 2022.11.29., Available at: <http://www.kbmaeil.com/news/articleView.html?idxno=944313>.
- [6] 김원희, 권영옥, “산업별 지속가능경영 전략 고찰: ESG 보고서와 뉴스 기사를 중심으로”, *지능정보연구*, 제29권, 제3호, 2023, pp. 287-316.
- [7] 남승현, *해외 인권과 환경 공급망 실사법 현황 및 시사점*, 국립외교원 외교안보연구소, 2022.
- [8] 대한상공회의소, *2023년 ESG 주요 현안과 정책 과제 조사*, 2023.
- [9] 박병립, “EU ESG 공급망 실사 법제화하면 반도체 등 주요업종 영향권...110여개 기업 고위험군”, *이투데이*, 2022.3.31., Available at <https://www.etoday.co.kr/news/view/2119299>.
- [10] 박수빈, 이용규, “텍스트 마이닝을 활용한 금융업 세부 업종간 ESG 보고서 비교 분석”, *국가정책연구*, 제36권, 제1호, 2022, pp. 31-56.
- [11] 산업통상자원부, 2023년 8월 수출입 동향, 2023.
- [12] 연다인, 박가연, 김희용, “텍스트 마이닝 기반

- 사용자 경험 분석 및 관리: 스마트 스피커 사례”, *경영정보학연구*, 제22권, 제2호, 2020, pp. 77-99.
- [13] 이동우, 민진영, “토픽 모델링과 머신 러닝 방법을 이용한 온라인 C2C 중고거래 시장에서 서의 사기 탐지 연구”, *경영정보학연구*, 제23권, 제4호, 2021, pp. 45-67.
- [14] 이수열, 이준겸, “글로벌 공급망의 지속가능 공급사슬 관리와 성과: 베트남 수출중심 공급사를 대상으로 한 실증연구”, *대한경영학회지*, 제28권, 제2호, 2015, pp. 453-468.
- [15] 이지원, 이향숙, “SNA 기법을 활용한 물류산업 ESG 키워드 분석: 뉴스기사 및 지속가능경영보고서를 활용하여”, *무역학회지*, 제47권, 제2호, 2022, pp. 121-132.
- [16] 장혜진, 최윤정, “ESG환경에서의 공급망 실사법”, *환경법 연구*, 제45권, 제1호, 2023, pp. 255-282.
- [17] 전효정, 박병조, 김태성, “정보보호 직무 수행을 위해 필요한 지식 및 기술: 텍스트 마이닝을 이용한 구인광고와 NCS의 비교”, *경영정보학연구*, 제25권, 제3호, 2023, pp. 179-197.
- [18] 한가록, 문두철, 이재은, “EU 공급망 실사지침(안)의 주요내용과 한국 중소기업의 ESG 전략: 키워드 네트워크 분석을 중심으로”, *중소기업연구*, 제44권, 제4호, 2022, pp. 41-66.
- [19] 한국ESG기준원, *2022년 ESG 평가 및 등급 공표*, 2022.
- [20] 한국품질재단, “지속가능경영보고서(Sustainability Report, SR)”, 2023, Available at https://www.kfq.or.kr/_KR/_Verification/Verification_02.aspx.
- [21] 황준석, 장현숙, *2023년 주목해야 할 EU 주요 환경규제와 대응전략*, 한국무역협회, 2023.
- [22] Blei, D. M., A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation”, *Journal of Machine Learning Research*, Vol.3, No.1, 2003, pp. 993-1022.
- [23] CDP, “The A List 2022”, 2023, Available at <https://www.cdp.net/en/companies/companies-scores>.
- [24] Clatworthy, M. A., “Differential patterns of textual characteristics and company performance in the chairman’s statement”, *Accounting, Auditing & Accountability Journal*, Vol.19, No.4, 2006, pp. 493-511.
- [25] Corporate Knights, “100 most sustainable companies of 2023 still flourishing in tumultuous times”, 2023, Available at <https://www.corporateknights.com/rankings/global-100-rankings/2023-global-100-rankings/2023-global-100-most-sustainable-companies/>.
- [26] Ergün, E. and Y. K. Usluel, “An analysis of density and degree-centrality according to the social networking structure formed in an online learning environment”, *Educational Technology & Society*, Vol.19, No.4, 2016, pp. 34-46.
- [27] European Commission, Proposal for a directive of the European Parliament and of the council on Coporate Sustainability Due Diligence and amending directive (EU) 2019/1937, 2022.
- [28] Griffiths, T. L. and M. Steyvers, “Finding scientific topics”, *PNAS(Proceedings of the National Academy of Sciences of the United States of America)*, Vol. 101, suppl. 1, 2004, pp. 5228-5235.
- [29] Kim, Y. S. and Y. J. Choi, “Textmining analysis on conceptualizing embedded topics in EU ESG discussion”, *Journal of Asia-Pacific Studies*, Vol. 29, No.4, 2022, pp. 191-223.
- [30] MSCI, “ESG Ratings”, 2023, Available at <https://www.msci.com/zh/esg-ratings>.
- [31] OECD, *OECD Due Diligence Guidance for Responsible Business Conduct*, 2018, pp. 1-3.
- [32] Okcu, E. U., “The 2021 EU trade policy review: shift in trade strategy and new initiatives”, *Ankara Avrupa Çalışmaları Dergisi*, Vol.21, No.2, 2022, pp. 497-518.
- [33] Abadi, M. S. N. and A. K. Kheyr, “A comparative

- analysis of mission and vision statements of top fortune companies and Iranian companies in food industry using text mining and clustering”, *Journal of Future Studies Management*, Vol.28, No.111, 2018, pp. 29-52.
- [34] Sharma, A., K. Kaushik, P. Awasthy, and A. Gawande, “Leveraging text mining for trend analysis and comparison of sustainability reports: Evidence from Fortune 500 companies”, *American Business Review*, Vol.25, No.2, 2022, pp. 416-438.
- [35] Tanaka, R. and K. Tsuda, “Differences on topics between the awarded and non-awarded integrated reports using text mining”, *Procedia Computer Science*, Vol.207, 2022, pp. 1446-1453.
- [36] UNGP, UN Guiding Principles Reporting Framework with implementation guidance, 2017, pp. 1-113.

A Study on the Countmeasures of the Korean Pharmaceutical/ Bio Industry to the EU Corporate Sustainability Due Diligence Directive, by using Text Mining

Sori Kim* · Joonhak Ki**

Abstract

In February 2022, the EU announced a draft of the EU Corporate Sustainability Due Diligence Directive requiring due diligence and disclosure of information on environmental and human rights risks in corporate supply chains.

This study evaluated the ability of 13 Korean pharmaceutical/bio companies to respond to the EU's demand for due diligence in the supply chain and compared it to 13 globally leading pharmaceutical/bio companies which are considered good in environmental and human rights risk management.

For comparative analysis, text mining analysis was performed using R. Basic word frequency and concurrent words were analyzed and topic modeling was performed by applying Latent Dirichlet Allocation.

As a result of the analysis, it was found that compared to advanced companies, domestic pharmaceutical and bio companies lack negative issue reporting and identification systems and supply chain due diligence implementation processes, and require advancement of data management for environmental and human rights information disclosure.

Accordingly, domestic pharmaceutical and bio companies need to prepare differentiated support measures to systematically identify and reduce risks in the supply chain of small and medium-sized businesses beyond simply providing financial support.

It is also desirable for the government to provide policy support by mandating Korea's own supply chain environment and human rights due diligence system, along with support for strengthening the ability to respond to due diligence of domestic pharmaceutical and bio companies, such as expert consulting and financial support.

Keywords: *ESG Management, Text Mining, EU Supply Chain Due Diligence, Pharmaceutical/
Bio Industry, Comparative Analysis*

* Corresponding Author, Researcher, ESG Management Center at Korean Standards Association

** Professor, The Department of Climate Environment and Energy Studies at Sookmyung Women's University

○ 저 자 소 개 ○



김 소 리 (srkim1208@ksa.or.kr)

숙명여자대학교 일반대학원 기후환경에너지학과에서 석사학위를 취득하였으며 현재 한국표준협회 ESG경영센터 연구원으로 재직 중이다. 주요 관심 분야는 ESG 경영전략 및 공시, 공급망 ESG 관리, R을 이용한 텍스트 마이닝 등이다.



기 준 학 (jhki@sookmyung.ac.kr)

서울대학교 자연과학대학원에서 박사학위를 취득하고 한화그룹 환경연구소장을 역임하였으며 현재 숙명여자대학교 산학협력단 산학협력교수로 재직 중이다. 주요 연구 분야는 ESG경영, 빅데이터경영, 탄소중립정책 등이다.

논문접수일 : 2023년 09월 14일

1차 수정일 : 2023년 10월 26일

게재확정일 : 2023년 12월 27일

2차 수정일 : 2023년 12월 05일